Statistical modeling with stochastic processes

Alexandre Bouchard-Côté

Lecture 6, Wednesday March 16

Program for today

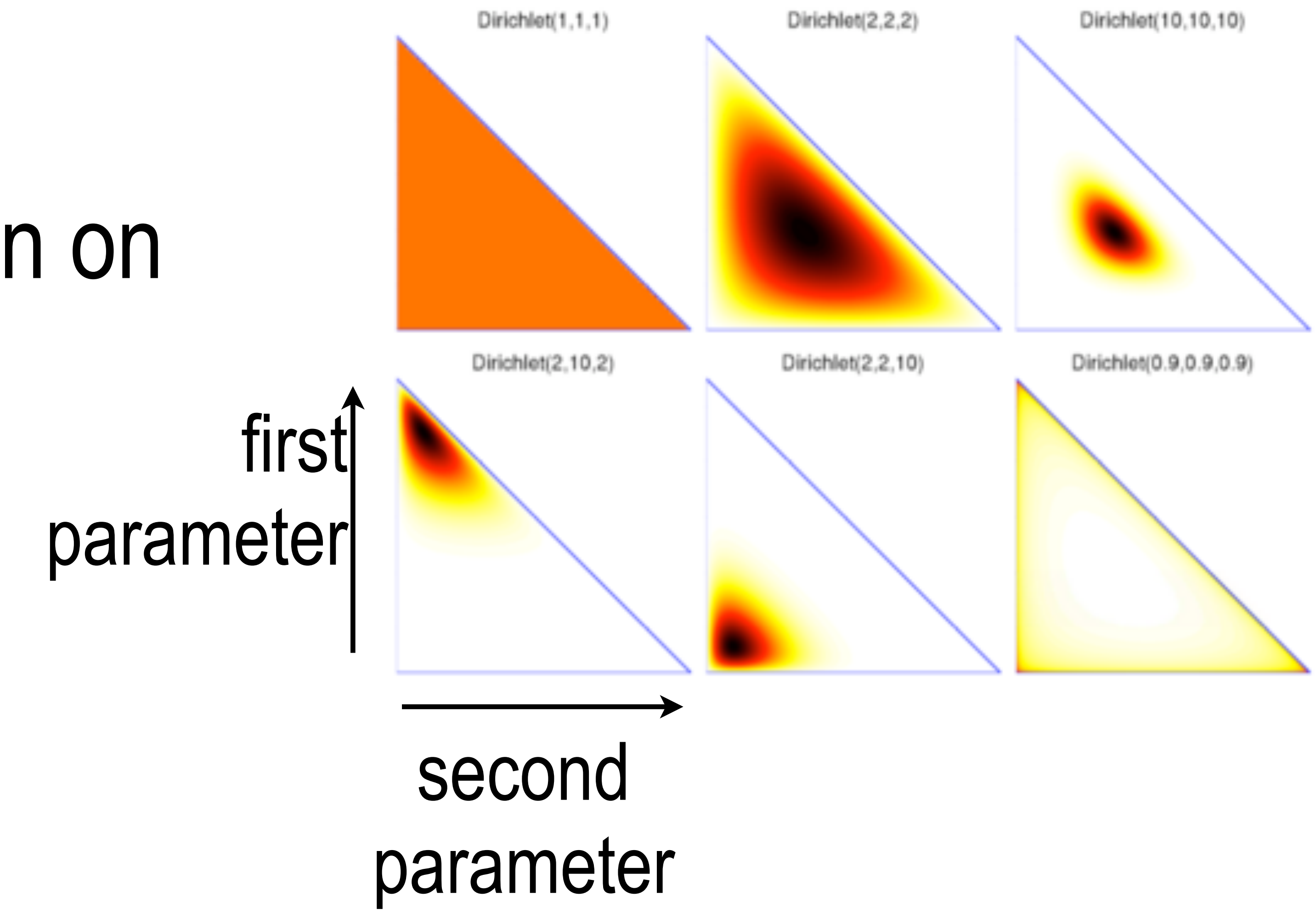
- Introduction to Bayesian non-parametrics
    - The Dirichlet Process: Theoretical foundations
    - Basic properties: posterior conjugacy, predictive distribution, etc
    - Chinese Restaurant, Polya Urn, etc.
- Basic probabilistic inference
    - Collapsed sampler
    - Slice sampler

Review

(Finite) Dirichlet distribution

Distribution on the parameters of categorical/multinomial distributions

Equivalent: distribution on the simplex

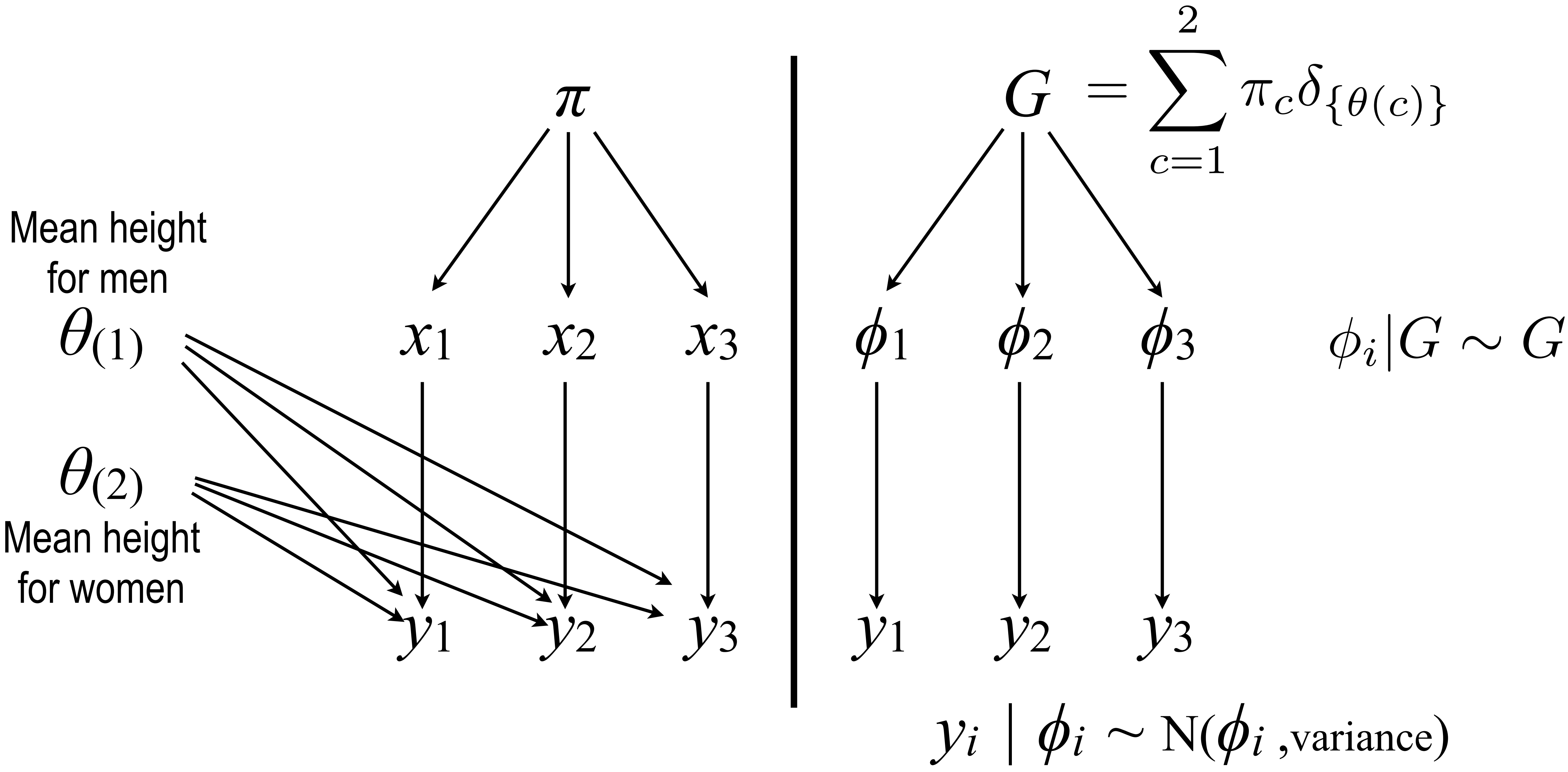


Density: for $\alpha_i > 0$

$$\frac{1}{Z(\alpha)}\prod_{i=1}^{K} x_i^{\alpha_i-1}\cdot\mathbf{1}\left[\sum_i x_i = 1, x_i \geq 0\right]$$

Equivalent notation

Mixture model: (UBC student height with 2 components)
say we have only 3 observations



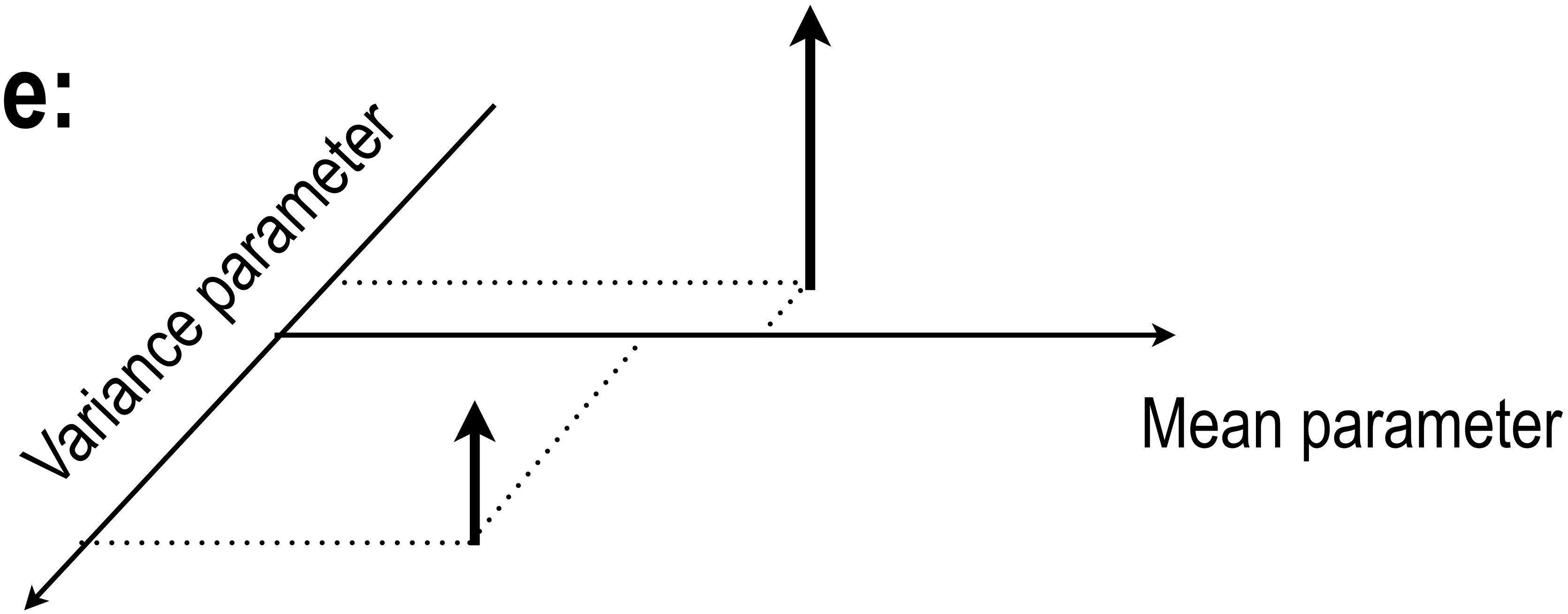
$$G = \sum_{c=1}^{2} \pi_c \delta_{\{\theta(c)\}}$$

$$\phi_i | G \sim G$$

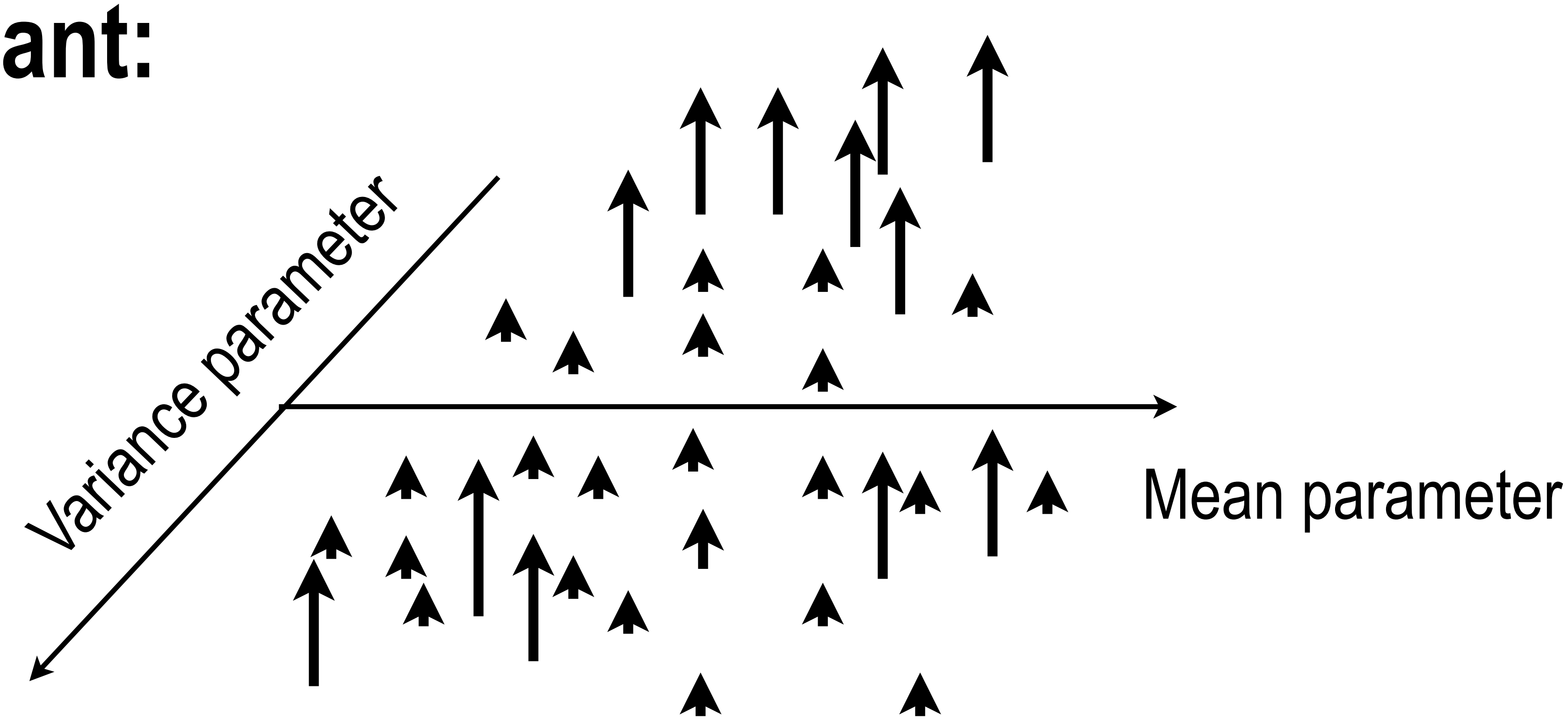
$$y_i \mid \phi_i \sim \mathrm{N}(\phi_i \text{ ,variance})$$

Samples from G

What we have:



What we want:



Definition: Dirichlet Process

Let G_0 be a distribution on a sample space Ω (the base distribution) α_0 be a positive real number (the concentration), and $(A_1, ..., A_\mathrm{k})$ be a partition of Ω. We say

$$G \sim \mathrm{DP}(\alpha_0, G_0)$$

i.e., G is a Dirichlet Process, if

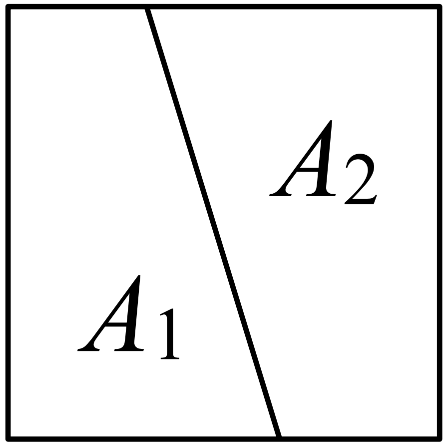
$$(G(A_1), \ldots, G(A_k)) \sim \mathrm{Dir}(\alpha_0 G_0(A_1), \ldots, \alpha_0 G_0(A_k))$$

for all partitions and all k.

Does this make sense/exists?

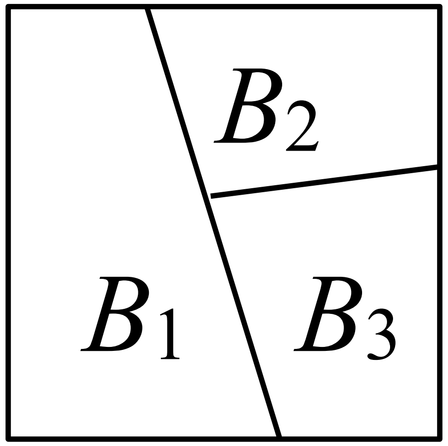
Kolmogorov consistency: check the marginals are consistent under marginalization

In this case: check that the marginals are consistent when refining partitions



$(G(A_1), G(A_2))$

(U_1, U_2)



$(G(B_1), G(B_2), G(B_3))$

(V_1, V_2, V_3)

Constructive argument

Claim: the random probability distribution constructed below is the Dirichlet process with base distribution G_0 and concentration α_0

$$\beta_j \overset{\text{iid}}{\sim} \text{Beta}(1, \alpha_0)$$

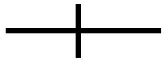
$$\theta_c \overset{\text{iid}}{\sim} G_0$$

We will denote this distribution over π by GEM(α_0)

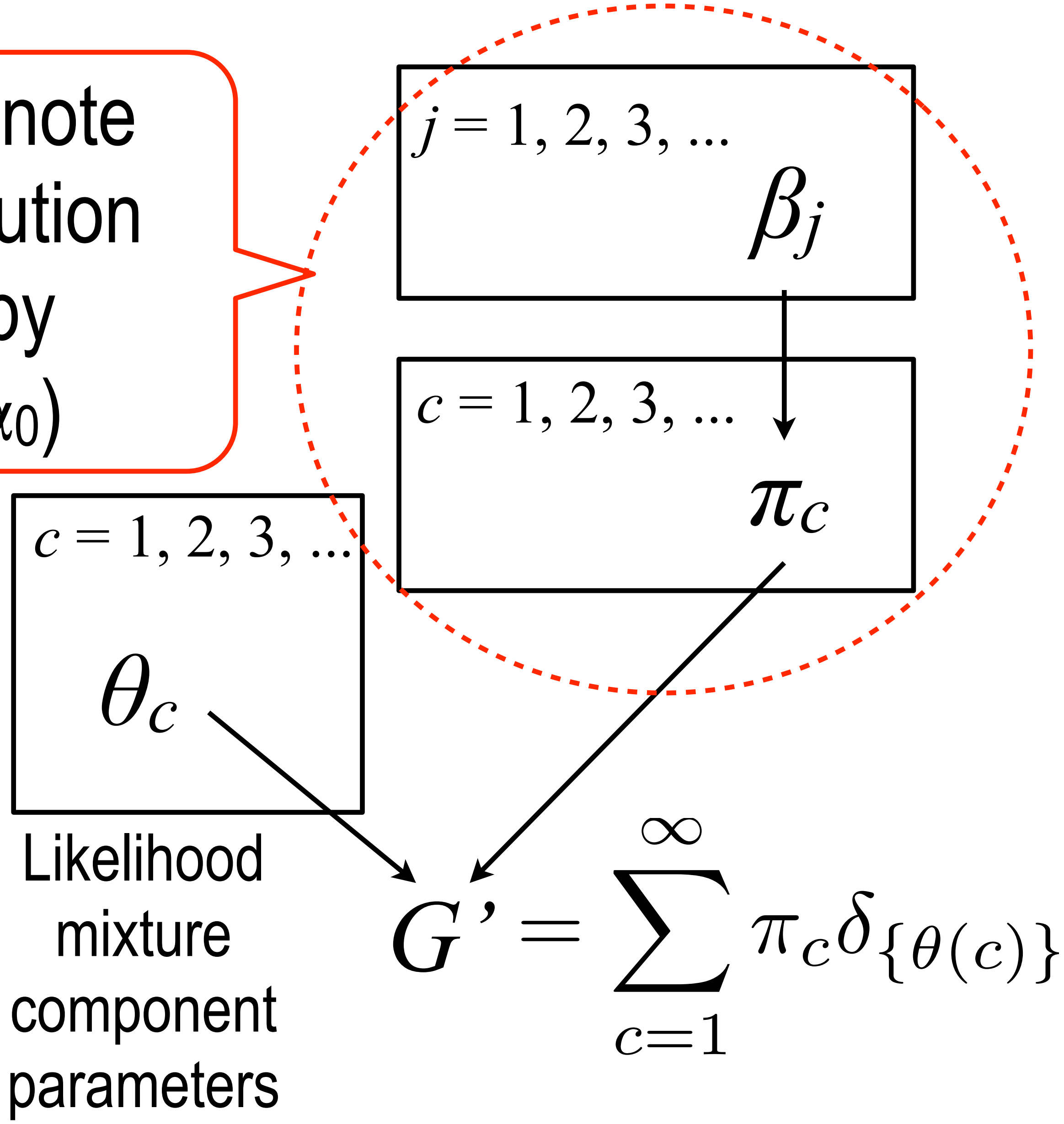
Start with a stick of length 1, and break a segment of length β_1 for π_1, keep the rest

$$\pi_1 = \beta_1$$

At step c, if the length of the stick remaining is L, set:

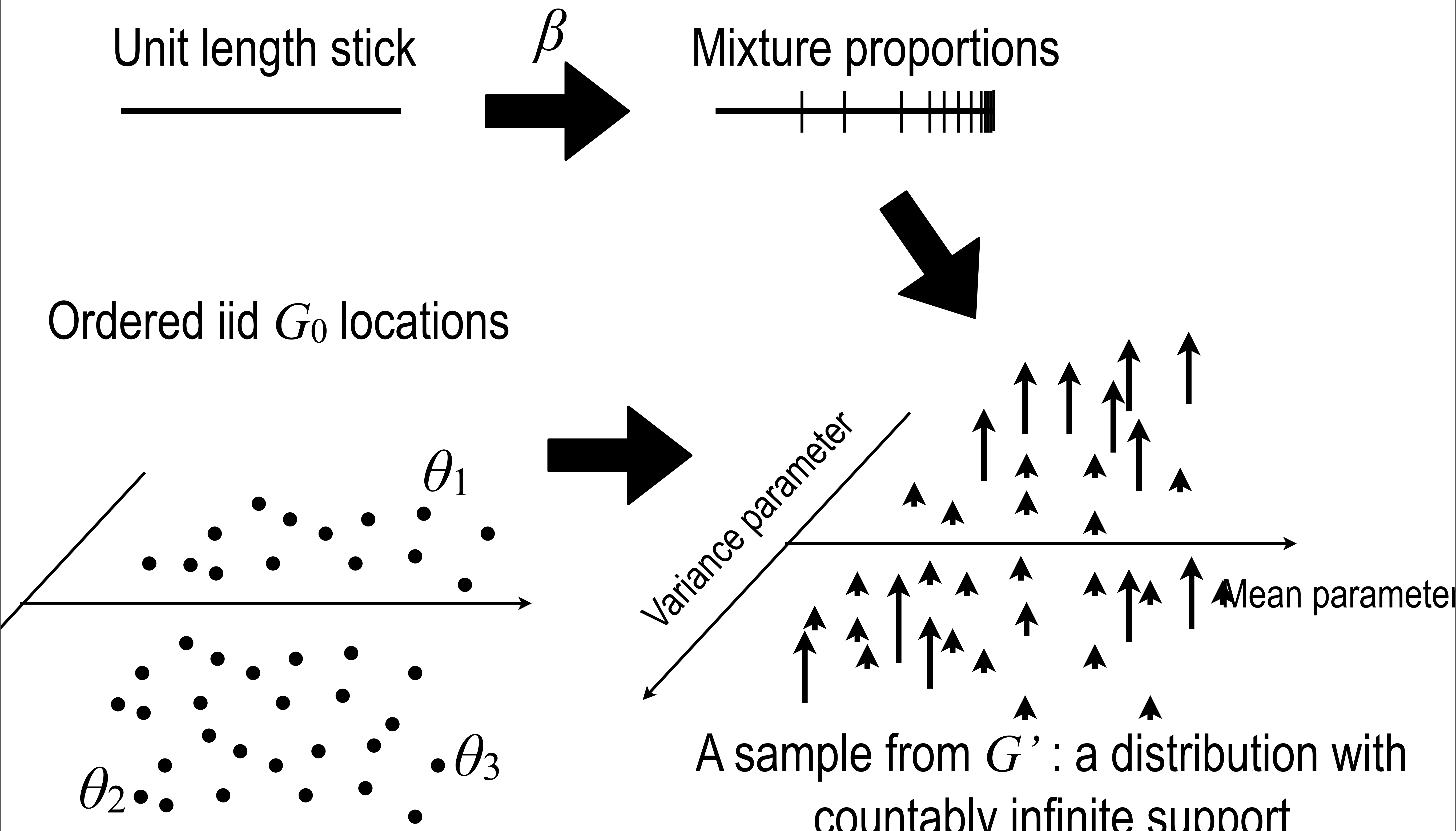


$$\pi_c = \beta_c L = \beta_c \prod_{j:j<c} (1 - \beta_j)$$



$$G' = \sum_{c=1}^{\infty} \pi_c \delta_{\{\theta(c)\}}$$

Samples from G



Unit length stick

β

Mixture proportions

Ordered iid G_0 locations

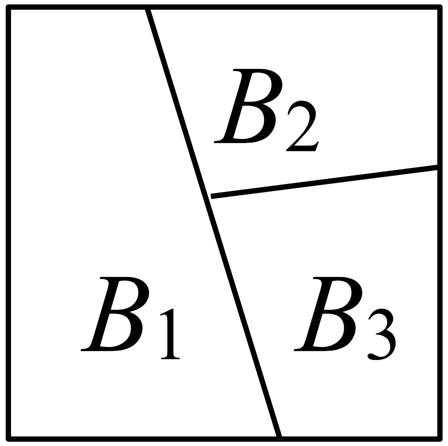
A sample from G' : a distribution with countably infinite support

Back to the proof that $G = G'$ in distribution

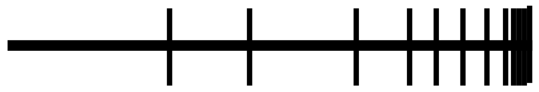
Reference: 'A constructive definition of Dirichlet Priors' (1994) Jayaram Sethuraman.

Goal: showing two definitions are equivalent

Kolmogorov consistency



Stick-breaking construction



Strategy: showing that for all partitions (A_1, ..., A_k), the constructed process G' has finite Dirichlet marginals

$$(G'(A_1), \ldots, G'(A_k)) \sim \mathrm{Dir}(\alpha_0 G_0(A_1), \ldots, \alpha_0 G_0(A_k))$$

Key observation: 'self-similarity'

Definitions: $$G' = f(\beta, \theta) = \sum_{c=1}^{\infty} \pi_c \delta_{\{\theta(c)\}}$$

$$\beta^* = (\beta_1, \beta_2, \ldots)^* = (\beta_2, \beta_3, \ldots)$$

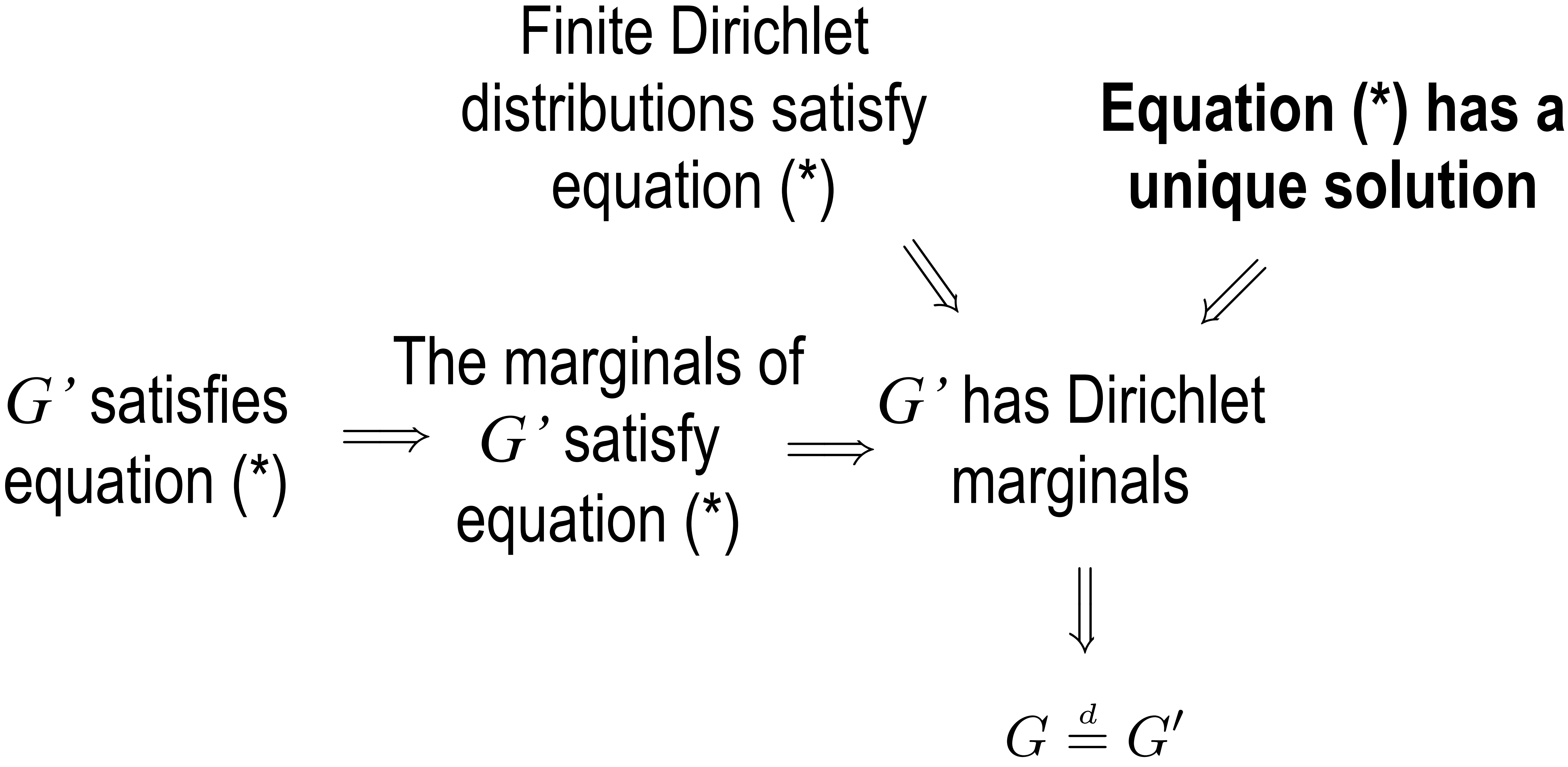
Observation: $$G' = \pi_1 \delta_{\{\theta(1)\}} + (1 - \pi_1) f(\beta^*, \theta^*)$$

$$= \pi_1 \delta_{\{\theta(1)\}} + (1 - \pi_1) G''$$ for $G' \overset{d}{=} G''$

Notation: $$G' \overset{st}{=} \pi_1 \delta_{\{\theta(1)\}} + (1 - \pi_1) G'$$ *

How we'll use it: we will show that if there is a distribution that satisfies this equation, it is unique; and that the finite Dirichlet distribution satisfies it

Detailed plan



Finite Dirichlet distributions satisfy equation (*)

Equation (*) has a unique solution

G' satisfies equation (*) $\Longrightarrow$ The marginals of G' satisfy equation (*) $\Longrightarrow$ G' has Dirichlet marginals

$$\Downarrow$$

$$G \overset{d}{=} G'$$

Lemma: uniqueness of the solution of *

Notation:

$$\underbrace{G'}_{V} \underset{\overset{st}{=}}{=} \underbrace{\pi_1\delta_{\{\theta(1)\}}}_{U} + \underbrace{(1-\pi_1)}_{W}\underbrace{G''}_{V} \qquad \textbf{*}$$

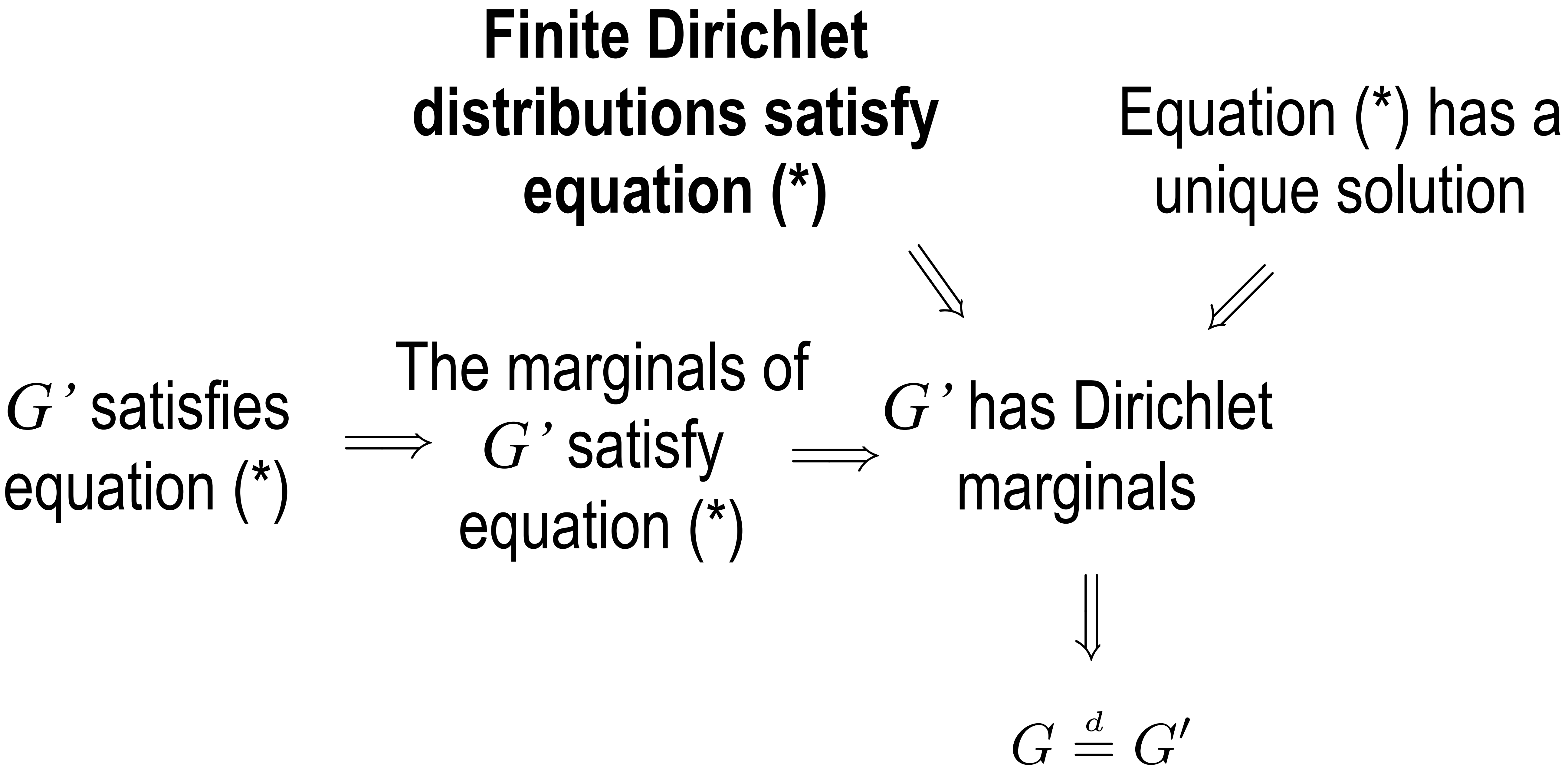
Properties we use:

- G'' is independent of (U, W)

- $P(0 < W < 1/2) > 0$

(Proof of uniqueness on the board)

Detailed plan



Finite Dirichlet distributions satisfy equation (*)

Equation (*) has a unique solution

G' satisfies equation (*) $\Longrightarrow$ The marginals of G' satisfy equation (*) $\Longrightarrow$ G' has Dirichlet marginals

$\Downarrow$

$$G \overset{d}{=} G'$$

Lemmas

Lemma 1. Let: $U \sim \mathrm{Dir}(\alpha_1, \dots, \alpha_k),$ $\alpha_0 = \sum_{i=1}^{k} \alpha_i$

$V \sim \mathrm{Dir}(\gamma_1, \dots, \gamma_k)$

$W \sim \mathrm{Beta}(\alpha_0, \gamma_0)$ all indep.

Then: $WU + (1 - W)V \sim \mathrm{Dir}(\alpha + \gamma)$

Proof: Gamma/neutral representation

Lemmas

Lemma 2. Let: $e_j = \text{a unit basis vector}$

$$\bar{\gamma}_j = \frac{\gamma_j}{\gamma_0}$$

Then: $$\sum_{j=1}^{k} \bar{\gamma}_j \mathrm{Dir}(\gamma + e_j) \sim \mathrm{Dir}(\gamma)$$

Proof: Exercise (next assignment)

Main proof: finite Dirichlet satisfies *

Goal: showing that:

$$V \sim \mathrm{Dir}(\gamma_1, \dots, \gamma_k)$$

satisfies equation (*) projected to the marginal of a finite partition

Steps:

1- Condition on the partition indicator X in which the first atom falls in

2- Sum over the possible values of X

Detailed plan

$$\begin{tikzcd}
 & \text{Finite Dirichlet distributions satisfy equation (*)} \arrow[dr, Rightarrow] & & \text{Equation (*) has a unique solution} \arrow[dl, Rightarrow] \\
G\text{' satisfies equation (*)} \arrow[r, Rightarrow] & \text{The marginals of } G\text{' satisfy equation (*)} \arrow[r, Rightarrow] & G\text{' has Dirichlet marginals} \arrow[d, Rightarrow] & \\
 & & G \overset{d}{=} G' &
\end{tikzcd}$$

Main properties of Dirichlet Processes

Moments

Let $G \sim \mathrm{DP}(\alpha_0, G_0)$ and A be a measurable set

Exercise: find the first and second moments of $G(A)$

(Derivation of the moments on the board)

Towards conjugacy

Let $G \sim \mathrm{DP}(\alpha_0, G_0)$ and $(A_1, ..., A_k)$ be a measurable partition. Let θ be a draw from G, i.e.: $\theta \mid G \sim G$

By multinomial-Dirichlet conjugacy, we have:

$$\big(G(A_1), \ldots, G(A_k)\big)\big|\theta \sim$$

$$\mathrm{Dir}\big(\alpha_0 G_0(A_1) + \delta_{\{\theta\}}(A_1), \ldots, \alpha_0 G_0(A_k) + \delta_{\{\theta\}}(A_k)\big)$$

Since this is true for all partitions, this means the posterior is a Dirichlet process as well!

Reference: 'The theory of statistics' (1995) Mark J. Schervish

Conjugacy

Found:

$$\big(G(A_1),\ldots,G(A_k)\big)\big|\theta \sim$$

$$\mathrm{Dir}\big(\alpha_0 G_0(A_1)+\delta_{\{\theta\}}(A_1),\ldots,\alpha_0 G_0(A_k)+\delta_{\{\theta\}}(A_k)\big)$$

Next: Identifying the new parameters α'_0 and G'_0 of the posterior distribution...

$$\alpha_0' = \sum_k \Big(\alpha_0 G_0(A_k) + \delta_{\{\theta\}}(A_k)\Big) = \alpha_0 + 1$$

$$G_0' = \frac{\alpha_0}{\alpha_0+1} G_0 + \frac{1}{\alpha_0+1}\delta_{\{\theta\}}$$

General formula

Suppose now we have several draws from G:

$$\big(G(A_1),\ldots,G(A_k)\big)\,\big|\,\theta_1,\ldots,\theta_n \sim$$
$$\mathrm{Dir}\big(\alpha_0 G_0(A_1)+n_1,\ldots,\alpha_0 G_0(A_k)+n_k\big)$$

where: $n_j = \sum_{i=1}^{n} \delta_{\{\theta_i\}}(A_j)$

Therefore the posterior parameters are:

$$\alpha_0' = \alpha_0 + n$$

$$G_0' = \frac{\alpha_0}{\alpha_0+n} G_0 + \frac{1}{\alpha_0+n}\sum_{i=1}^{n}\delta_{\{\theta_i\}}$$

Predictive distribution

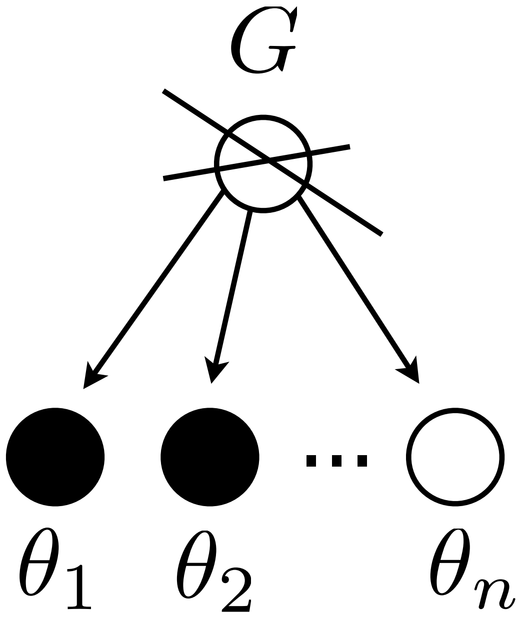
Predictive distribution: $\theta_{n+1}|\theta_1, \ldots, \theta_n$

Motivation for marginalizing *G*: posterior inference using MCMC on a *finite* state space

Let A be a measurable set, $G \sim \mathrm{DP}(G_0, \alpha_0)$

$\theta_1, \ldots, \theta_{n+1}|G \sim G$

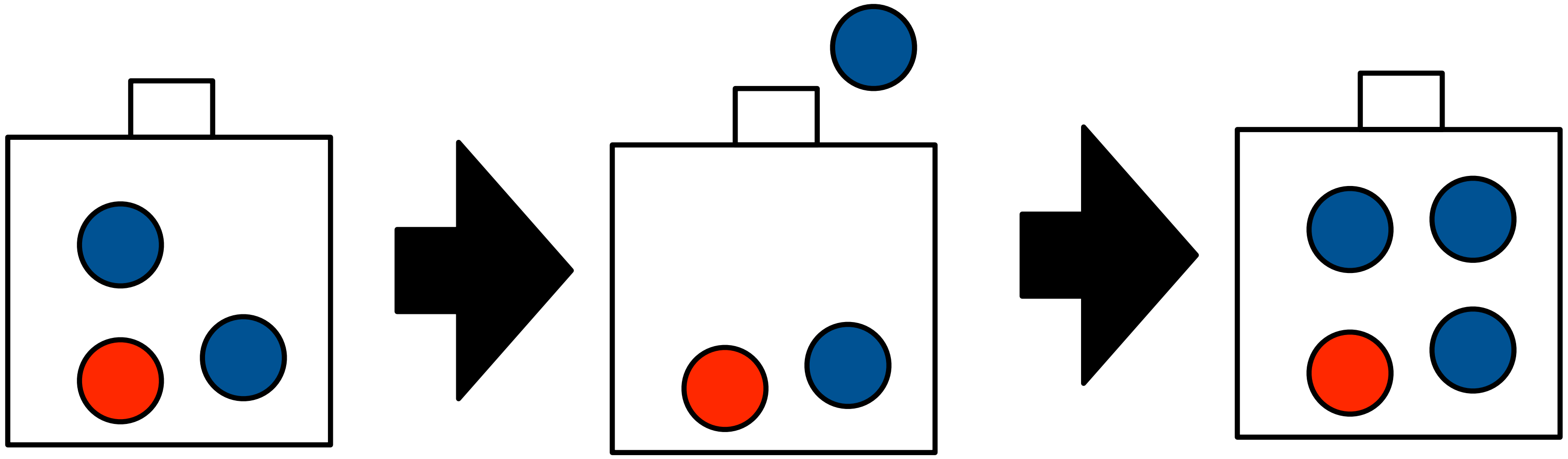
(Derivation of the predictive dist. on the board)



Application: Pólya Urn

Thought experiment: Consider an urn, with initially R red marbles and B blue marbles.

At each step, draw one marble at random, and put it back in the urn after adding another one of the same color



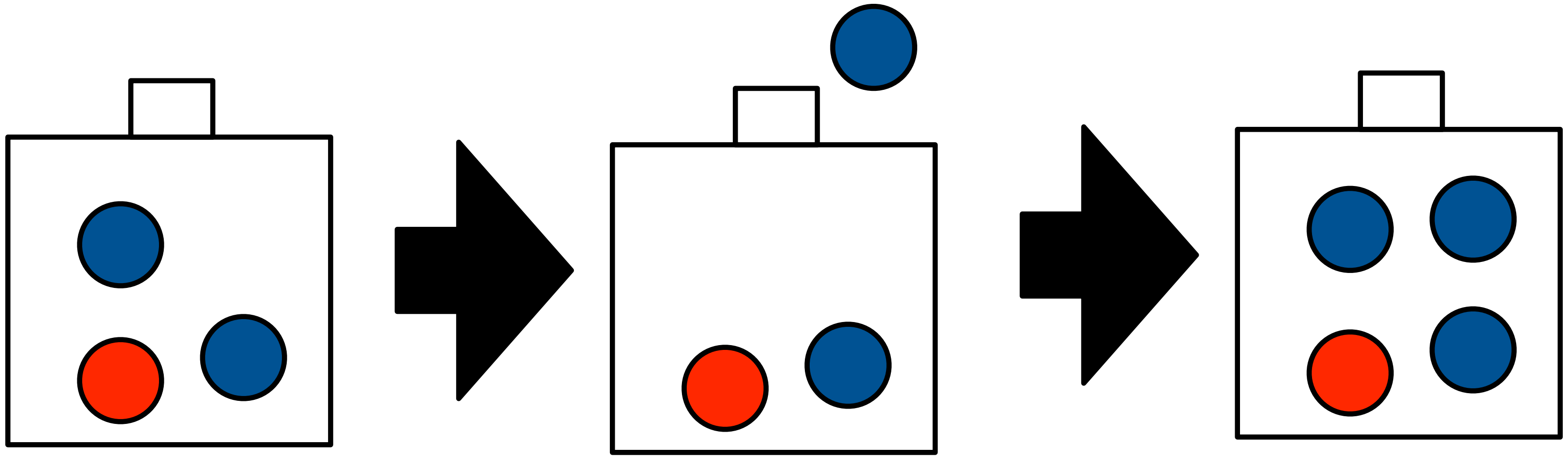
Question: does this process converge to a certain red:blue ratio? What is this ratio?

Application: Pólya Urn

Thought experiment: Consider an urn, with initially R red marbles and B blue marbles.

Question: does this process converge to a certain red:blue ratio? What is this ratio?

Hint: Let $\alpha_0 = R + B$ $\quad G_0 = \mathrm{Bern}\left(\frac{R}{B+R}\right)$



(Solution described on the board)

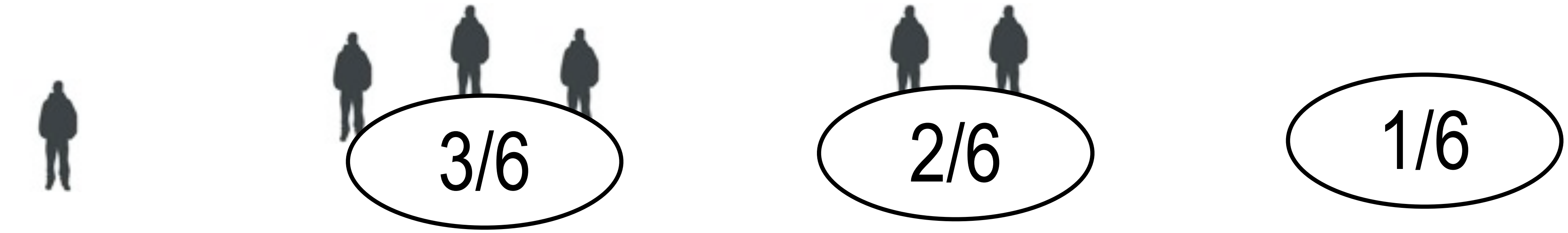
Chinese Restaurant Process (CRP)

Idea: Instead of colors sharing colors, think about customers sharing tables in an infinite restaurant

Initialization: The first customer sits in the first empty table.

Iterate: If n **customers** are already sitting in the restaurant, the next customer starts a new table with probability α_0 / $(\alpha_0 + n)$; otherwise the customer joins an existing table with probability proportional to the number of people already at the table

E.g.:



$\alpha_0 = 1$

New Join table 1? Join table 2? Start emtpy?