Statistical modeling with stochastic processes

Alexandre Bouchard-Côté

Lecture 11, Monday April 4

Program for today

- Beta, Poisson and Gamma processes
- DDP and sequence memoizer

Pitman-Yor process

Pitman-Yor process: Start with the CRP, and boost the probability of table creation while preserving exchangeability

This has the same normalization as the DP: $\alpha_0 + n$

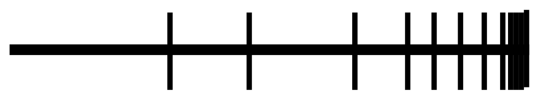
New customer

...

$\propto n_1 - d$	$\propto n_t - d$	$\propto \alpha_0 + t\,d$
Join table #1, with already n_1 people sitting there	Join table # t, with already n_t people sitting there	Create a new table

Discount: $d \in [0, 1]$

PY: stick breaking construction



Dirichlet process: defined $G = f(\beta, \theta)$ for an iid sequence of $\theta_i \sim G_0$ and:

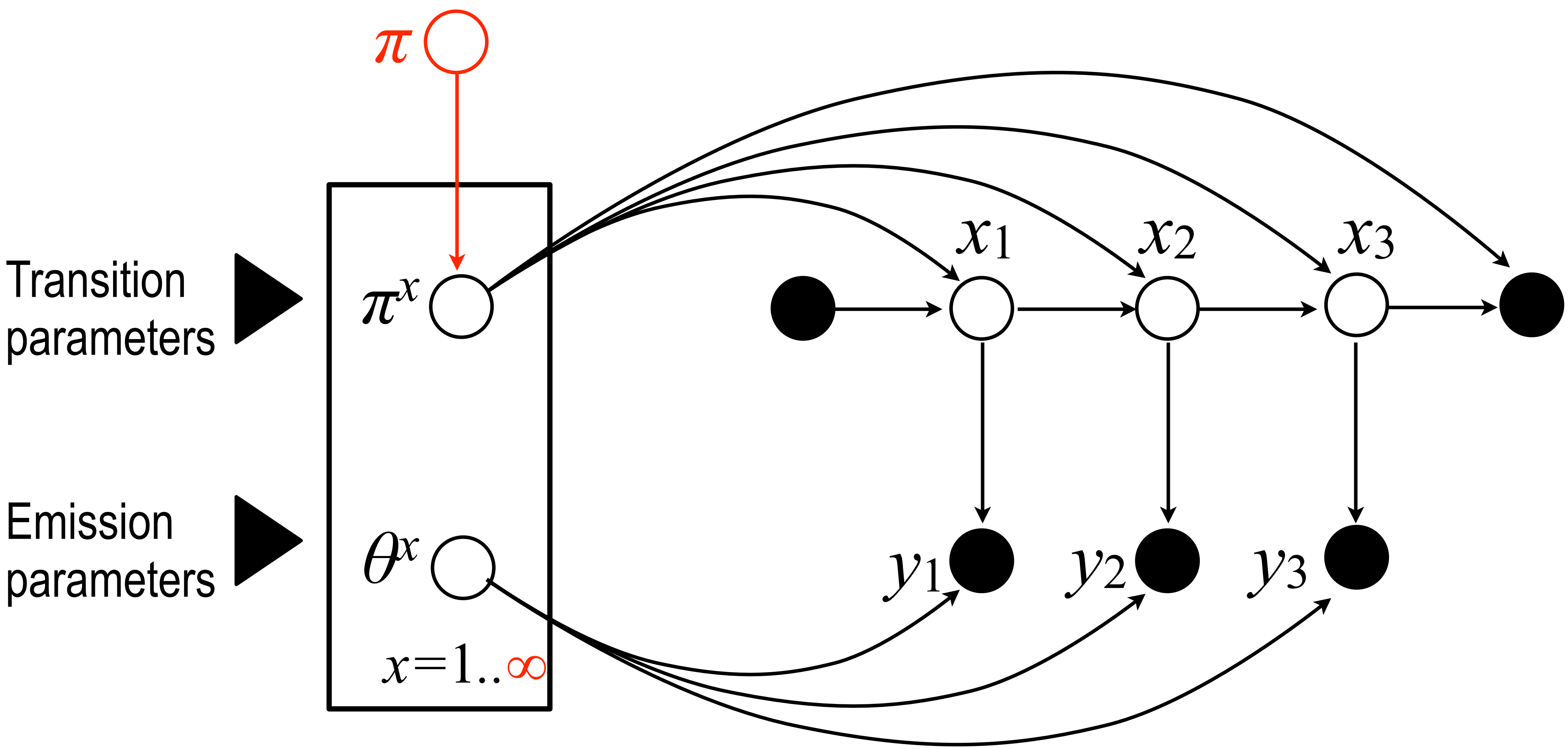
$$\beta_i \sim \mathrm{Beta}(1, \alpha_0),$$

Pitman-Yor: Same but now beta's are not identically dist.:

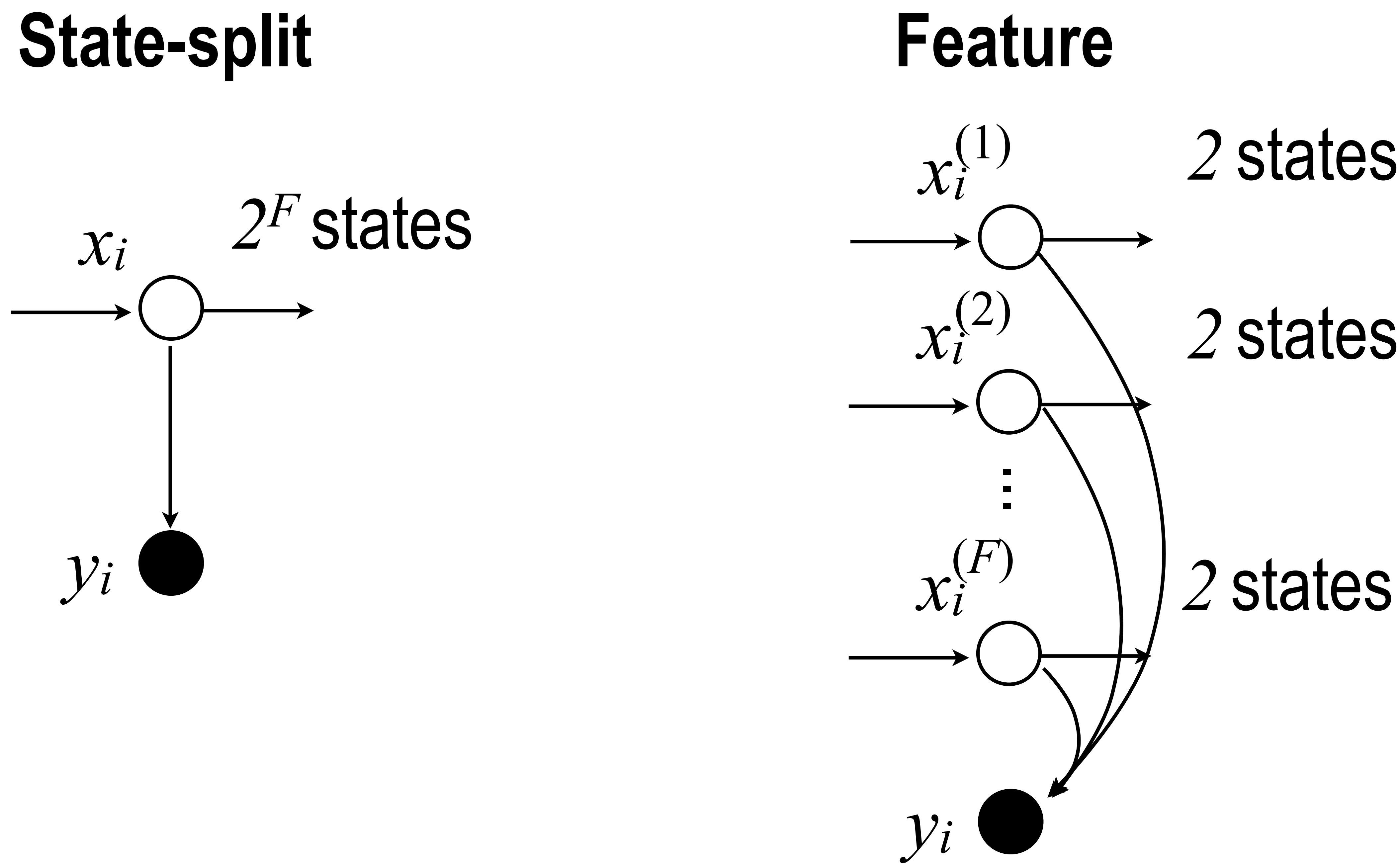
$$\beta_i \sim \mathrm{Beta}(1 - d, \alpha_0 + i\, d)$$

The infinite HMM

Infinite HMMs:



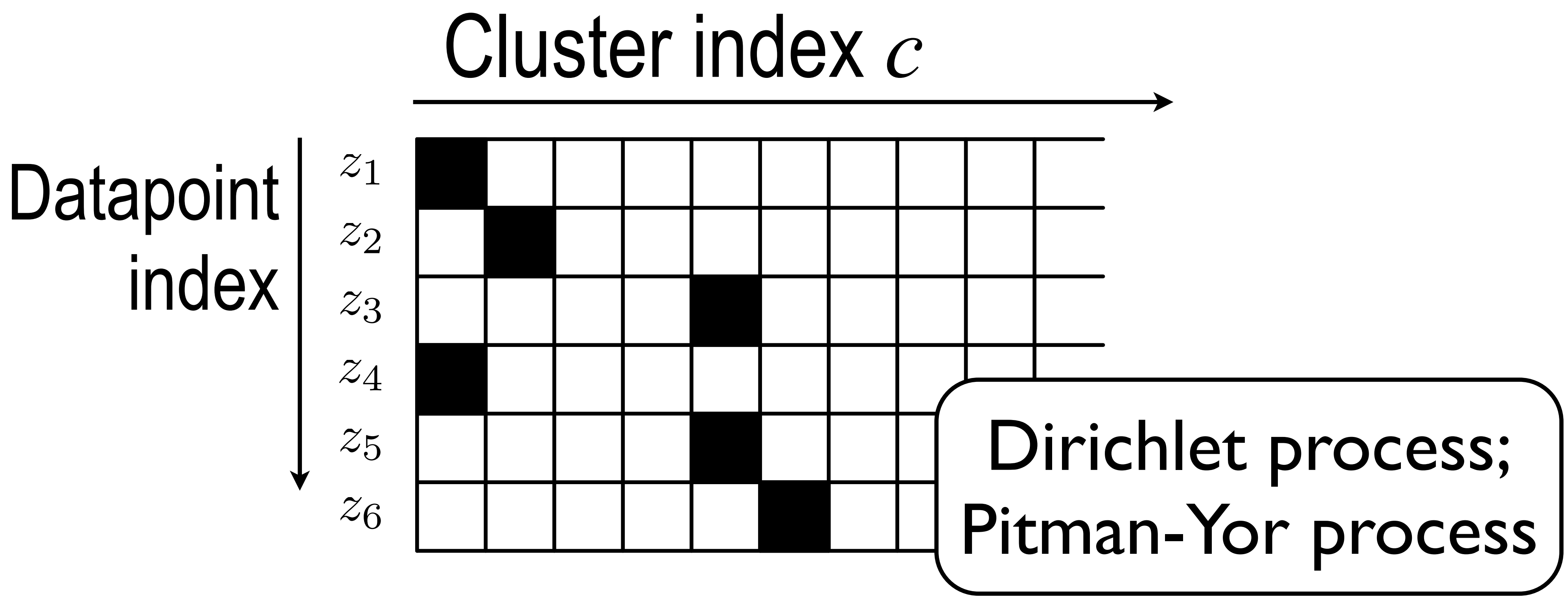
Feature based representations



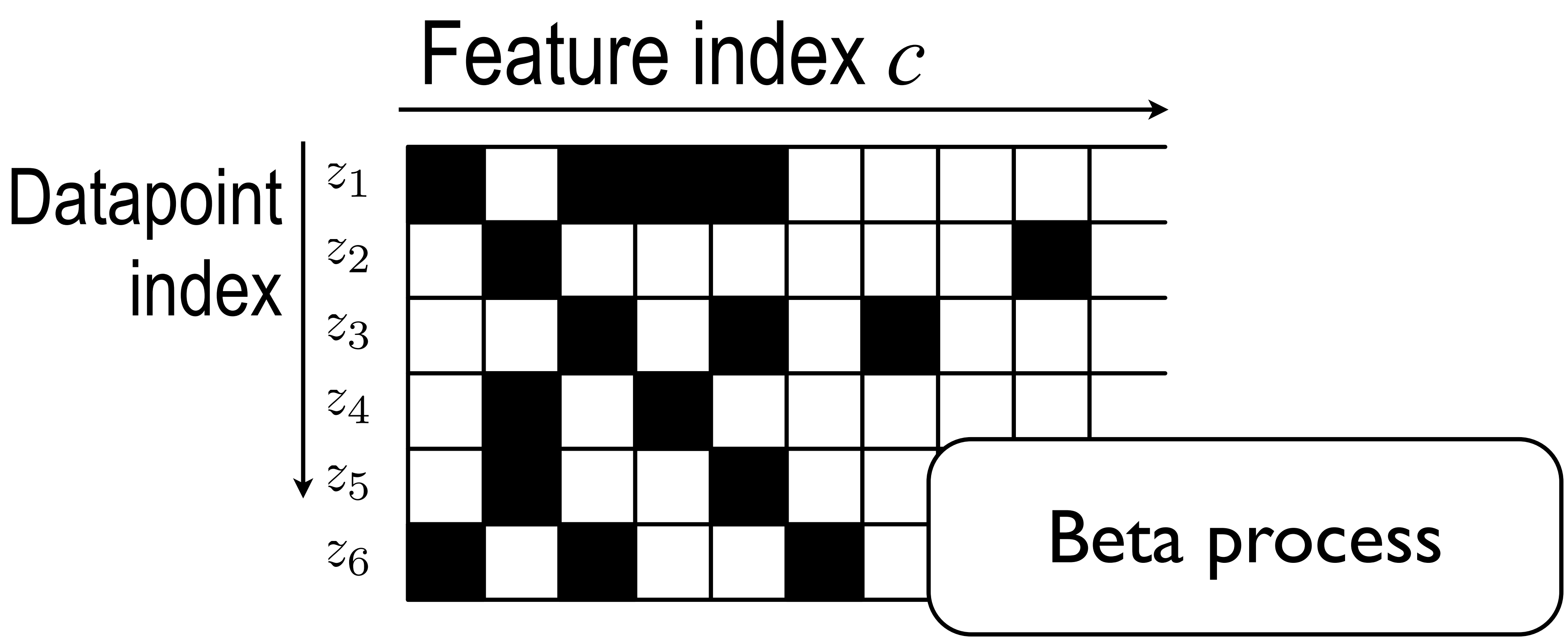
How many features? Will see soon a solution: Beta process

Beta process

Mixture indicator priors:

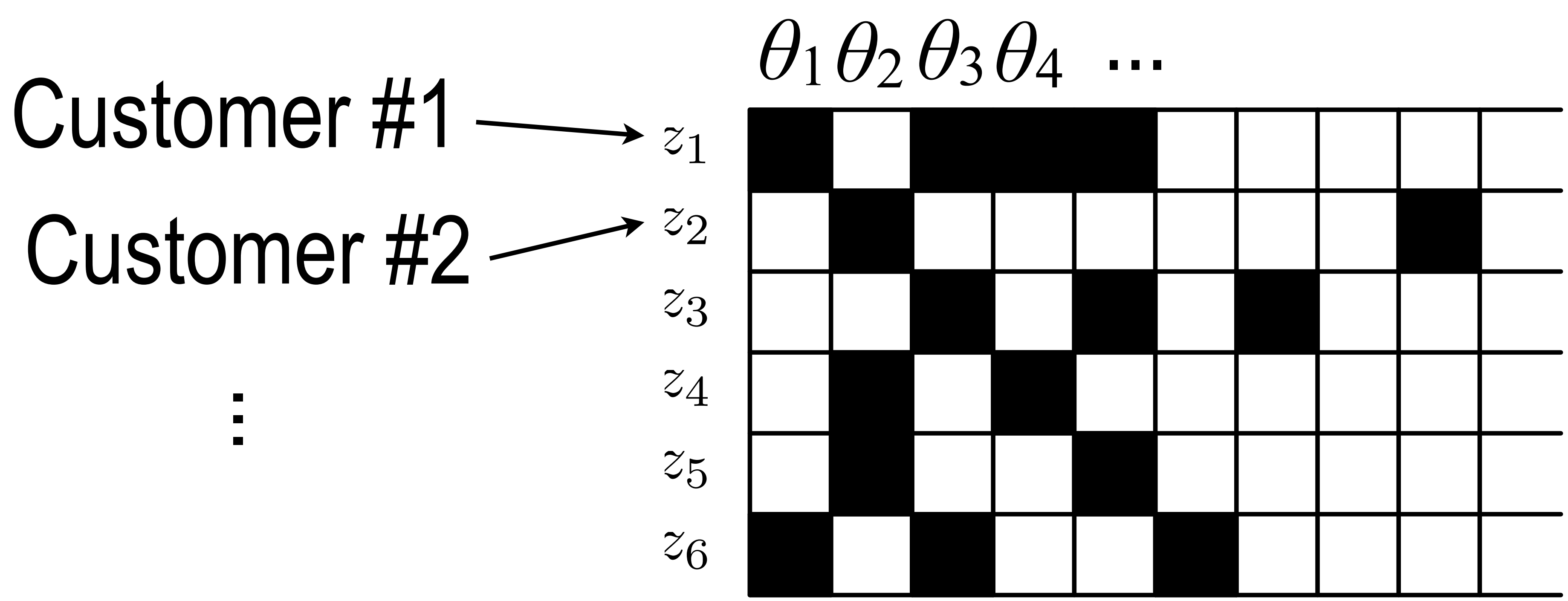


Feature indicator priors:



Predictive distribution: restaurant metaphor

Instead of a sit-down restaurant, think of a buffet with an infinite sequence of dishes θ_i sampled by customers

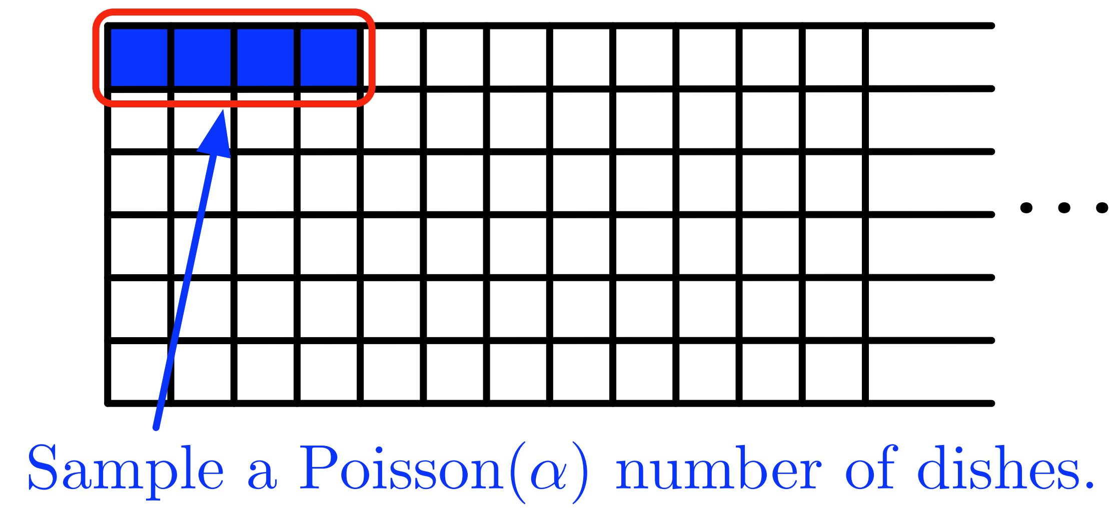


Obvious: order of the columns not important/exchangeable (because the θ_i's will be generated iid)

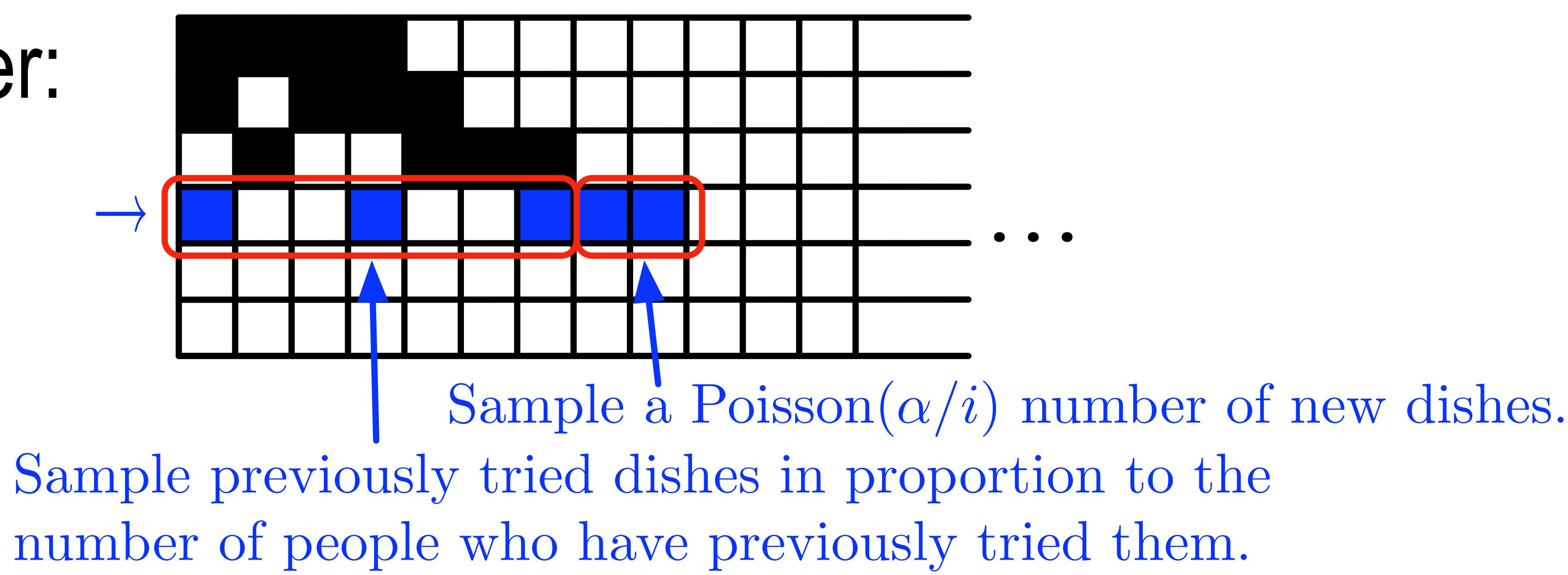
Less obvious: how to make the order of the rows exchangeable

Predictive distribution: restaurant metaphor

First customer:



Fourth customer:



(Example on the board)

Slide from Kurt Miller

Beta process: stick breaking representation

Interpretation of the sequence of sticks $(\pi_j)_{j=1..\infty}$

π_j is the prior probability of picking row j

Consequence: the sticks no longer sum to one!

Construction (will come back to it later):

Beta process:

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

$$\pi_k = \prod_{l=1}^{k} (1 - \beta_l)$$

Cf.: Dirichlet process

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

$$\pi_k = \beta_k \prod_{l=1}^{k-1} (1 - \beta_l)$$

Poisson processes

Poisson processes

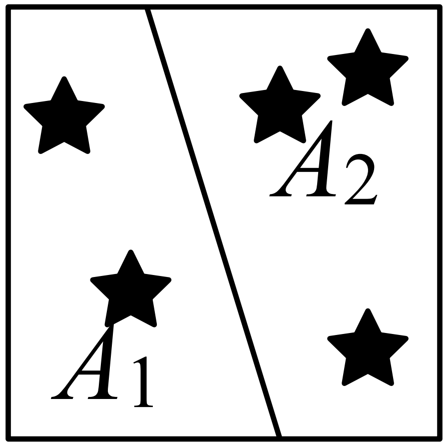
Another random discrete measure, but unnormalized:

Let P_0 be a distribution on a sample space Ω (the base distribution) and $(A_1, ..., A_k)$ be a partition of Ω. We say

$$P \sim \mathrm{PP}(P_0)$$

i.e., P is a Poisson Process, if

$$P(A_1) \overset{\text{ind.}}{\sim} \mathrm{Poi}(P_0(A_1))$$



for all partitions and all k.

Cf: 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.

Consistency/existence

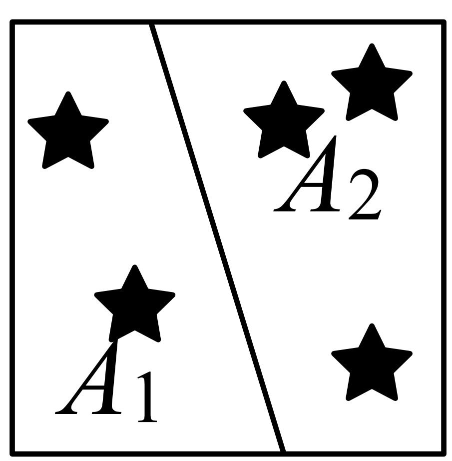
Let P_0 be a distribution on a sample space Ω (the base distribution) and $(A_1, ..., A_k)$ be a partition of Ω. We say

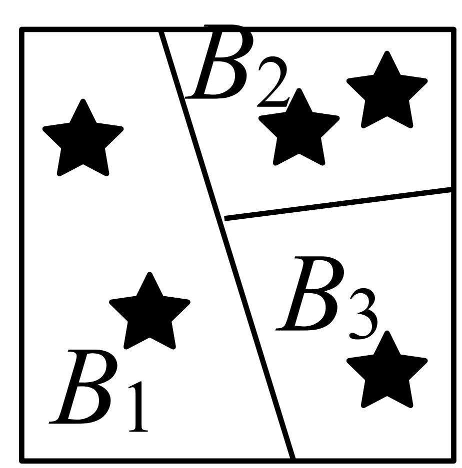
$$P \sim \mathrm{PP}(P_0)$$

i.e., P is a Poisson Process, if

$$P(A_1) \overset{\text{ind.}}{\sim} \mathrm{Poi}(P_0(A_1))$$

for all partitions and all k.





Campbell's theorem

Assume P_0 is a probability measure, f is bounded, and $P \sim \mathrm{PP}(P_0)$.

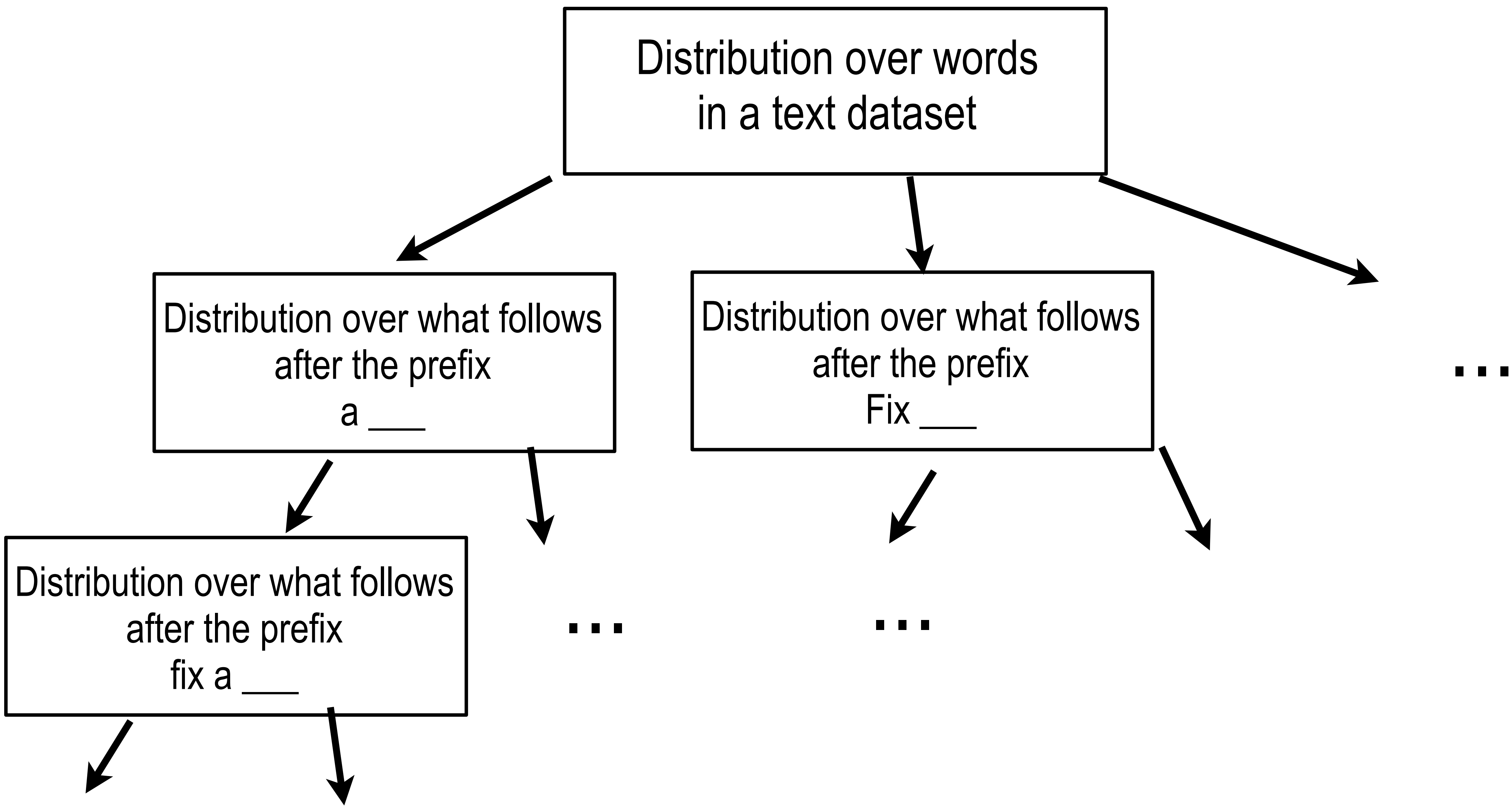
Let also: $$\Sigma = \sum_{X \in P} f(X)$$

Then: $$\mathbb{E}\left[e^{it\Sigma}\right] = \exp\left\{\int_{\Omega} (e^{itf(x)} - 1) P_0(\,\mathrm{d}x)\right\}$$

Sequence memoizer

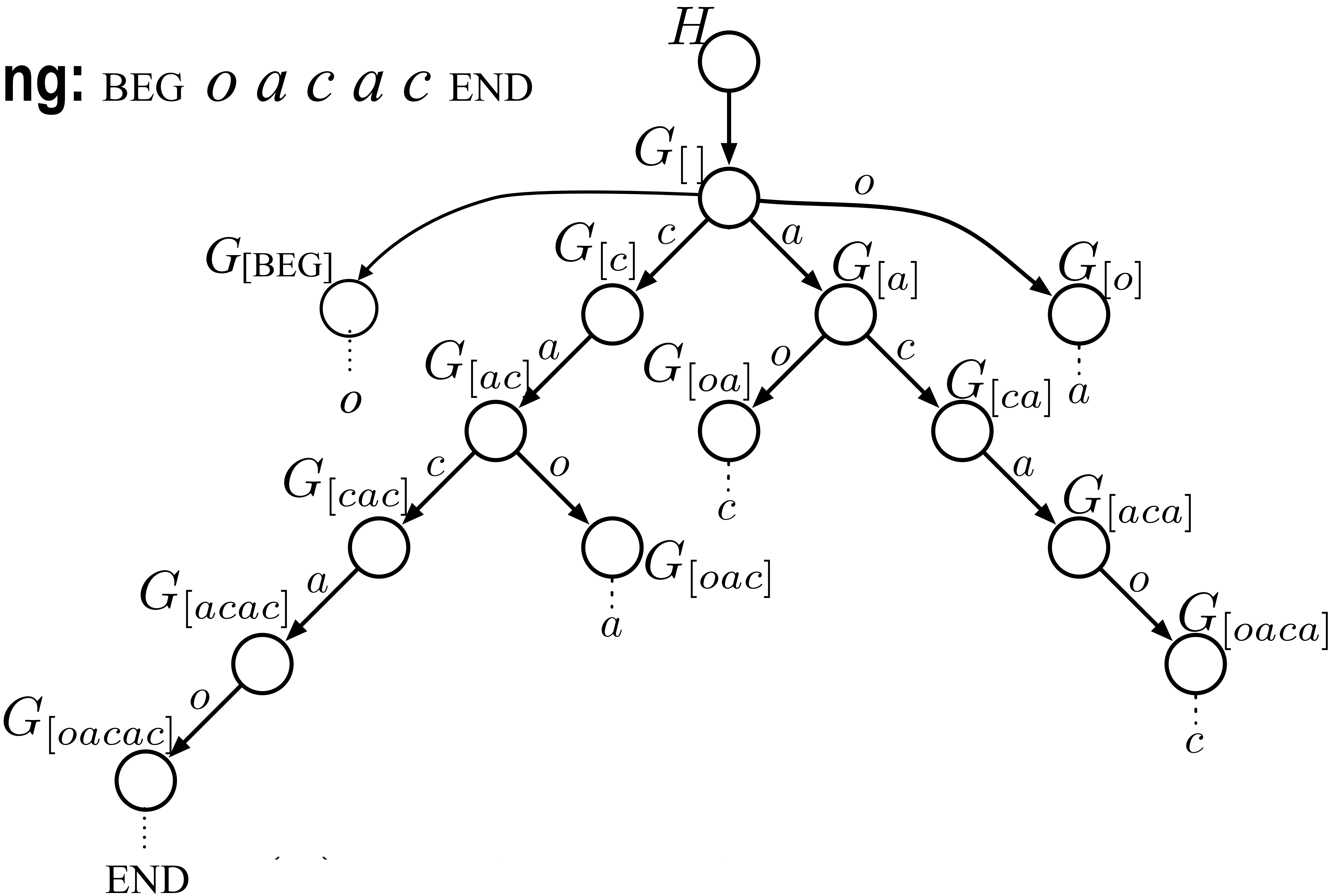
Back to hierarchical models

Hyper-prior over words---not specific to a prefix



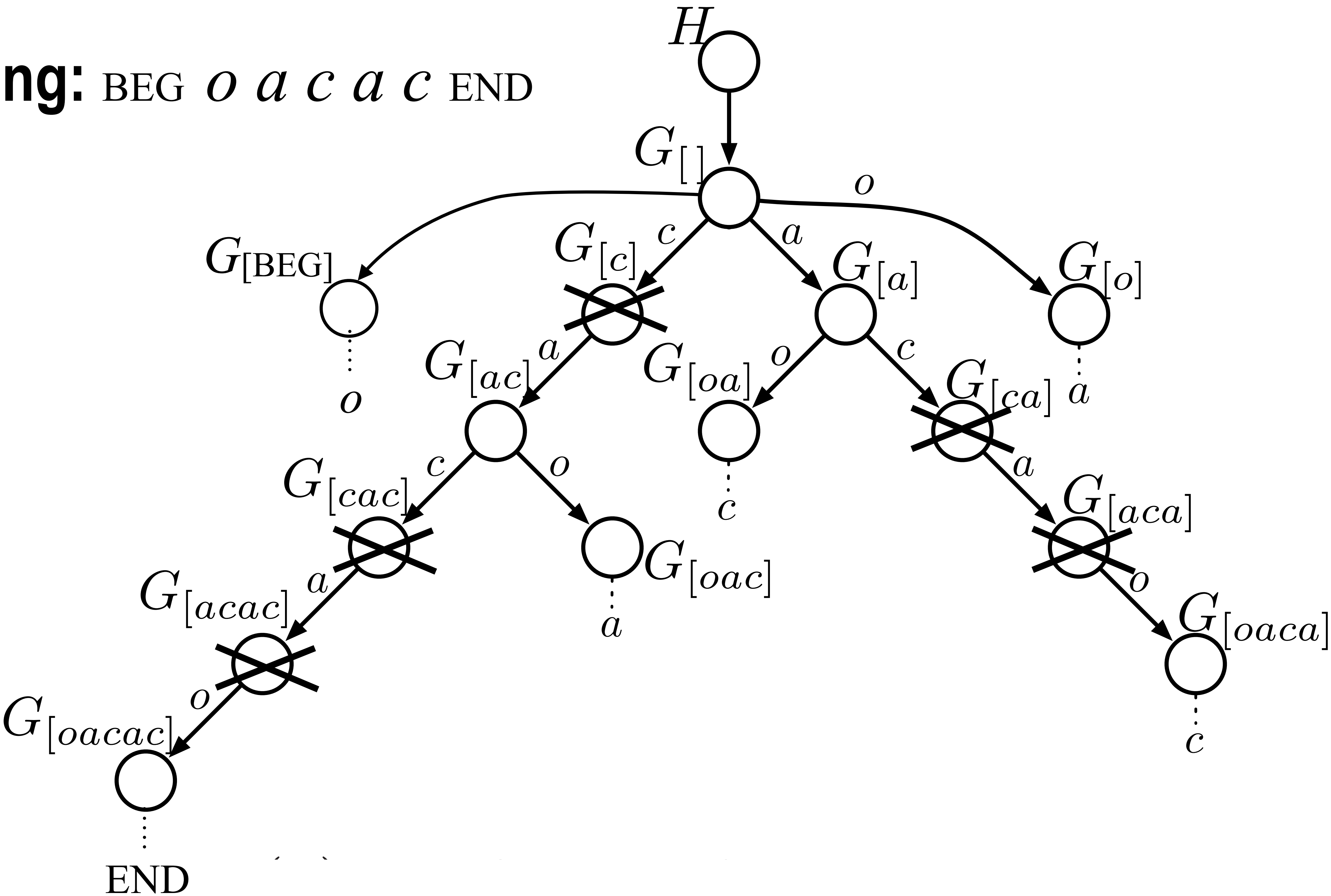
More elaborate example

Training: BEG $o\ a\ c\ a\ c$ END



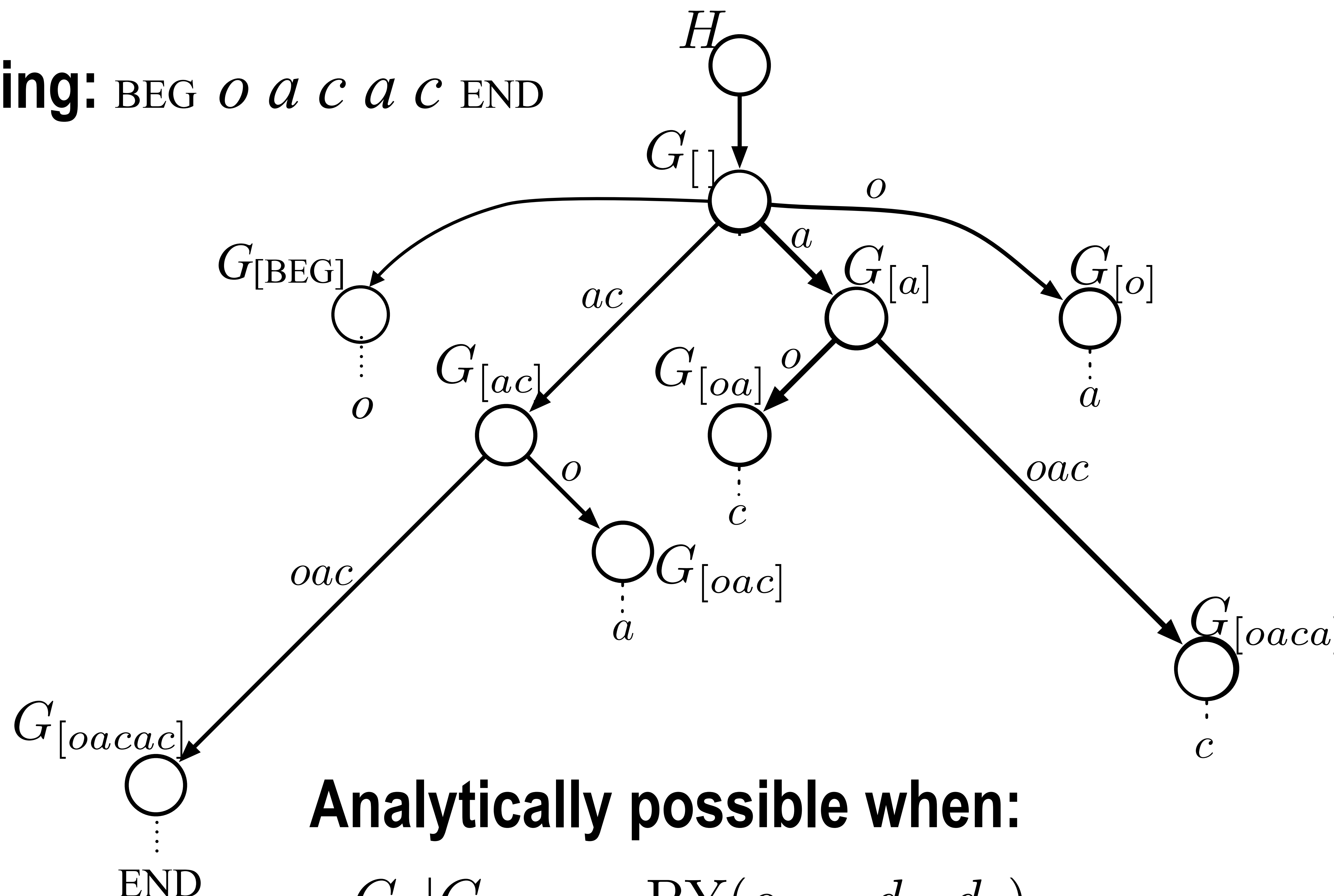
Marginalization

Training: BEG $o\ a\ c\ a\ c$ END



Analytic marginalization

Training: BEG $o\ a\ c\ a\ c$ END



Analytically possible when:

$$G_s | G_{\sigma(s)} \sim \mathrm{PY}(\alpha_{\sigma(s)} d_s, d_s)$$

Condition for analytic marginalization