CSE 5523: Lecture Notes 24
Infinite Mixture Models and Dirichlet Processes

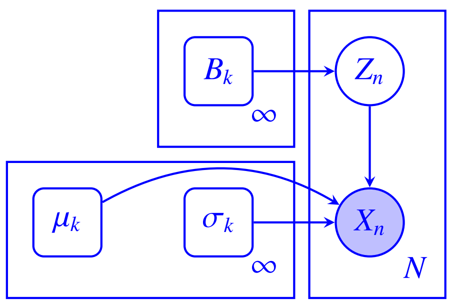
Contents

Mixture models give us nice clusters, but often we don't know how many of them we'll need.

24.1 Stick-breaking processes

Fortunately we can create arbitrarily large distributions using sequences of binomial choices.

Each choice gives the probability that we are 'there yet' as we count through the categories.



This is called a **stick-breaking process model** (a kind of **Dirichlet process model**).

$$\text{StickBreakingProcess}_{\alpha}(b_1, \ldots, b_\infty) = \prod_{k=1}^{\infty} \text{Beta}_{1,\alpha}(b_k)$$

The likelihood of any z_n is then a sequence of these 'are we there yet' binomial decisions:

$$\mathsf{P}_{b_1,\ldots,b_\infty}(z_n) = \prod_{n=1}^{N} \left(\prod_{k=1}^{z_n - 1} \text{Binomial}_{b_k}(0) \right) \text{Binomial}_{b_{z_n}}(1)$$

These can be sampled from Binomials, which can sample their parameters from Beta as needed.

We call this an '**infinite mixture model**', but it's bounded by the largest z_n sampled (value K).

Since the probabilities decline exponentially, the model will only ever need $\mathcal{O}(\log N)$ categories.

24.2 Conjugacy for stick-breaking processes

We can sample b_k's from the posterior (based on z_n values) using Beta-binomial conjugacy:

$$\mathsf{P}(b_{1..K} \mid z_{1..N}) = \frac{\mathsf{P}(b_{1..K}) \cdot \mathsf{P}(z_{1..N} \mid b_{1..K})}{\mathsf{P}(z_{1..N})} \qquad \text{Bayes' law}$$

$$
\begin{aligned}
&\propto \mathsf{P}(b_{1..K}) \cdot \mathsf{P}(z_{1..N} \mid b_{1..K}) && \text{def. proportion} \\
&= \overbrace{\prod_{k=1}^{K} \mathrm{Beta}_{1,\alpha}(b_k)}^{\text{prior}} \overbrace{\prod_{n=1}^{N} \left(\prod_{k=1}^{z_n-1} \mathrm{Binomial}_{b_k}(0) \right) \mathrm{Binomial}_{b_{z_n}}(1)}^{\text{likelihood}} && \text{substitution} \\
&= \left(\prod_{k=1}^{K} \frac{\Gamma(1+\alpha)}{\Gamma(1)\,\Gamma(\alpha)} (b_k)^0 \, (1-b_k)^{\alpha-1} \right) \prod_{n=1}^{N} \left(\prod_{k=1}^{z_n-1} 1-b_k \right) b_{z_n} && \text{def. Beta, Binom} \\
&\propto \left(\prod_{k=1}^{K} (b_k)^0 \, (1-b_k)^{\alpha-1} \right) \prod_{n=1}^{N} \left(\prod_{k=1}^{z_n-1} 1-b_k \right) b_{z_n} && \text{def. proportion} \\
&= \left(\prod_{k=1}^{K} (b_k)^0 \, (1-b_k)^{\alpha-1} \right) \prod_{k=1}^{K} \left(\prod_{j=1}^{k-1} 1-b_j \right) (b_k)^{\sum_n [\![z_n=k]\!]} && \text{prod. of exps} \\
&= \prod_{k=1}^{K} (b_k)^{0+\sum_n [\![z_n=k]\!]} \, (1-b_k)^{\alpha-1+\sum_n [\![z_n>k]\!]} && \text{prod. of exps} \\
&\propto \prod_{k=1}^{K} \mathrm{Beta}_{1+\sum_n [\![z_n=k]\!],\, \alpha+\sum_n [\![z_n>k]\!]}(b_k) && \text{def. Beta}
\end{aligned}
$$

What if we need more binomials than previous iterations? Just sample more with $\sum_n [\![z_n=k]\!] = 0$.

There are also variants of this (Chinese restaurant process) that marginalize out the b's.