Announcements  Tentative upcoming workload schedule:

- Competition 1: Dataset already released. Instructions out today-ish? Tentative due date Wed April 22, 11:59pm.
- A3: Released relative soon-ish, probably due somewhere in the weeks of the 20th or the 27th.
- Competition 2: Released maybe around the weeks of the 20th or 27th, due Mon May 11th, 4:30pm.

Outline:

1. Review mixture of multinomials notation (model, notation)
2. New setting: "longitudinal" data (analogy w/ documents), where same person can have different types?
   a. Choice of multinomials:
      - from fixed set $\rightarrow$ distribution over multinomials: the Dirichlet

\[ (\text{For handout}) \]

Last time: data $X$ has rows like $x_1 = (3, 0, 6, \ldots)$

- A generative story of multinomials
  - For a given customer $x_1$ (corresponding to the $i$th row), $x_1 = (x_{11}, x_{12}, \ldots, x_{1d})$
  - They pick among given customer types w/ prob

\[ \pi(1), \pi(2), \ldots, \pi(d) \]

- A customer type is represented by the parameters of a multinomial $\phi_j$
  - $\phi_j[1]$ (prob of picking a pepsi)
  - $\phi_j[2]$ (prob of picking a coke)

\[ \sum_{j=1}^{d} \phi_j[i] = 1, \quad \text{all } \phi_j[i] \geq 0. \]

We let $x_i$ be the type that customer $i$ picked.

$x_i$ then picks their purchases according to the $\phi_j$ they picked.
I. Clicker question: notation review  What is the (possibly unnormalized) probability of a single \( x_t \) according to our model, assuming we knew the hidden \( \pi \), \( \phi_j \)s, and \( c_t \)s?

\[
\begin{align*}
(A) \quad & \pi[c_t] \frac{m!}{x_t[1]!x_t[2]! \cdots x_t[d]!} \prod_{\ell=1}^d \phi_{ct}[\ell] x_t[\ell] \\
(B) \quad & \pi[c_t] \frac{m!}{\phi[1]!\phi[2]! \cdots \phi[d]!} \prod_{\ell=1}^d \phi_{ct}[\ell] x_t[\ell] \\
(C) \quad & x_t[c_t] \frac{m!}{\phi[1]!\phi[2]! \cdots \phi[d]!} \prod_{\ell=1}^d \pi_{ct}[\ell] \phi_{ct}[\ell] \\
(D) \quad & \phi[c_t] \frac{m!}{x_t[1]!x_t[2]! \cdots x_t[d]!} \prod_{\ell=1}^d \pi_{ct}[\ell] x_t[\ell]
\end{align*}
\]

II. Clicker question: new likelihood  What is the likelihood of a single \( x_t \) under our second model (after mixture of multinomials, before full-blown each customer has different preference profile)?

\[
\begin{align*}
(A) \quad & \prod_{q=1}^m \phi[c_t[q]] \pi_{ct[q]}[x_t[q]] \\
(B) \quad & \prod_{q=1}^m \pi[c_t[q]] \phi_{ct[q]}[x_t[q]]
\end{align*}
\]

III. From Percy Liang and Dan Klein’s 2007 tutorial, Structured Bayesian Nonparametric Models with Variational Inference  (on next page)
Dirichlet distribution

A Dirichlet is specified by concentration parameters:

\[ \alpha = (\alpha_1, \ldots, \alpha_K), \ \alpha_i \geq 0 \]

Mean: \( \left( \frac{\alpha_1}{\sum_{i=1}^{K} \alpha_i}, \ldots, \frac{\alpha_K}{\sum_{i=1}^{K} \alpha_i} \right) \)

Variance: larger \( \alpha \)s \( \rightarrow \) smaller variance

A Dirichlet draw \( \phi \) is written \( \phi \sim \text{Dirichlet}(\alpha) \),

which means \( p(\phi | \alpha) \propto \phi_1^{\alpha_1-1} \cdots \phi_K^{\alpha_K-1} \)

The Dirichlet distribution assigns probability mass to multinomials that are likely to yield pseudocounts \( (\alpha_1 - 1, \ldots, \alpha_K - 1) \).

Mode: \( \left( \frac{\alpha_1 - 1}{\sum_{i=1}^{K} (\alpha_i - 1)}, \ldots, \frac{\alpha_K - 1}{\sum_{i=1}^{K} (\alpha_i - 1)} \right) \)

The full expression for the density of a Dirichlet is \( p(\phi | \alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} \phi_i^{\alpha_i-1}. \) Note that unlike the Gaussian, the mean and mode of the Dirichlet are distinct. This leads to a small discrepancy between concentration parameters and pseudocounts: concentration parameters \( \alpha \) correspond to pseudocounts \( \alpha - 1 \).

Draws from Dirichlet distributions

- Dirichlet(5, 5, 5)
- Dirichlet(1, 1, 1)
- Dirichlet(5, 10, 8)

A Dirichlet(1, 1, 1) is a uniform distribution over multinomial parameters. As the concentration parameters increase, the uncertainty over parameters decreases. Going in the other direction, concentration parameters near zero encourage sparsity, placing probability mass in the corners of the simplex. This sparsity property is the key to the Dirichlet process.