# Gibbs Sampling for Bayesian Mixture

Carl Edward Rasmussen

November 27th, 2022

# Key concepts

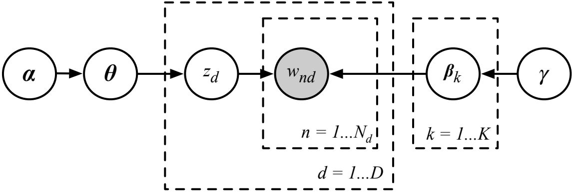
- General Bayesian mixture model
- We derive the Gibbs sampler
- Marginalize out mixing proportions: collapsed Gibbs sampler

# Bayesian document mixture model

Our mixture model has observations $w_d$ the words in document $d = 1, \ldots, D$.
The parameters are $\boldsymbol{\beta}_k$ and $\boldsymbol{\theta}$, and latent variables $z$.

The mixture model has $K$ components, so the parameters are $\boldsymbol{\beta}_k, k = 1, \ldots K$.
Each $\boldsymbol{\beta}_k$ is the parameter of a categorical over possible words, with prior $p(\boldsymbol{\beta})$.
The discrete latent variables $z_d, d = 1, \ldots D$ take on values $1, \ldots K$.

Note, that in this model the observations are (the word counts of) entire documents.




$$\begin{aligned}
\boldsymbol{\theta} &\sim \mathrm{Dir}(\alpha) \\
\boldsymbol{\beta}_k &\sim \mathrm{Dir}(\gamma) \\
z_d|\boldsymbol{\theta} &\sim \mathrm{Cat}(\text{`}) \\
w_{nd}|z_d, \boldsymbol{\beta} &\sim \mathrm{Cat}(\text{fi}_{z_d})
\end{aligned}$$

# Bayesian mixture model

The conditional likelihood is for each observation is

$$p(w_d|z_d = k, \boldsymbol{\beta}) \;=\; p(w_d|\beta_k) \;=\; p(w_d|\beta_{z_d}),$$

and the prior

$$p(\boldsymbol{\beta}_k) \;=\; \mathrm{Dir}(\gamma)$$

The categorical latent component assignment probability

$$p(z_d = k|\boldsymbol{\theta}) \;=\; \theta_k,$$

with a Dirichlet prior

$$p(\theta|\alpha) \;=\; \mathrm{Dir}(\alpha).$$

Therefore, the latent conditional posterior is

$$p(z_d = k|w_d, \boldsymbol{\theta}, \boldsymbol{\beta}) \;\propto p(z_d = k|\boldsymbol{\theta})p(w_d|z_d = k, \boldsymbol{\beta}) \;\propto\; \theta_k p(w_d|\beta_{z_d}),$$

which is just a discrete distribution with $K$ possible outcomes.

# Gibbs Sampling

Iteratively, alternately, sample the three types of variables:

Component parameters

$$p(\beta_k|w, z) \;\propto\; p(\beta_k) \prod_{d:z_d=k} p(w_d|\beta_k) \;\propto\; \text{Dir}(\gamma + c_{mk}),$$

with $c_{mk} = \sum_{d:z_d=k} c_{md}$. This is now a categorical model, the mixture aspect having been eliminated.

The posterior latent conditional allocations

$$p(z_d = k|w_d, \theta, \boldsymbol{\beta}) \;\propto \theta_k p(w_d|\beta_{z_d}),$$

are categorical and mixing proportions

$$p(\theta|z, \alpha) \;\propto\; p(\theta|\alpha)p(z|\theta) \;\propto\; \text{Dir}(c + \alpha).$$

where $c_k = \sum_{d:z_d=k} 1$ are the counts for mixture k.

# Collapsed Gibbs Sampler

The parameters are treated in the same way as before.

If we marginalize over $\theta$

$$\begin{aligned} p(z_d = k | z_{-d}, \alpha) &= \int p(z_d = k|\theta) p(\theta | z_{-d}, \alpha) d\theta \\ &= \int \theta_k p(\theta | z_{-d}, \alpha) d\theta = \frac{\alpha + c_{-d,k}}{\sum_{j=1}^{K} \alpha + c_{-d,j}}, \end{aligned}$$

where index $-d$ means *all except* $d$, and $c_k$ are counts;
we derived this result when discussing pseudo counts.

The collapsed Gibbs sampler for the latent assignements

$$p(z_d = k | w_d, z_{-d}, \boldsymbol{\beta}, \alpha) \propto p(w_d | \beta_k) \frac{\alpha + c_{-d,k}}{\sum_{j=1}^{K} \alpha + c_{-d,j}},$$

where now all the $z_d$ variables have become dependent (previously they were conditionally independent given $\theta$).

Notice, that the Gibbs sampler exhibits the *rich get richer* property.