CS340 Machine learning
Mixture priors
Conjugate priors

- We have used conjugate priors for computational convenience.
- But sometimes our prior beliefs cannot be expressed in this way.
- Eg coin is either 1/3 prob heads or 2/3 prob heads
Mixture of conjugate priors

- \( Z \in \{1, \ldots, K\} \) is a latent variable which specifies which prior mixture component to use.

\[
p(Z = k | \pi) = \pi_k
\]

\[
p(\theta | Z = k, \alpha) = p(\theta | \alpha_k)
\]

\[
p(x | \theta) = p(x | \theta)
\]

Example: Beta(\( \theta | 10, 20 \)) or Beta(\( \theta | 20, 10 \))
Posterior is also a mixture

\[
p(\theta | x) = \frac{p(x | \theta)p(\theta)}{p(x)}
\]

\[
= \frac{p(x | \theta) \sum_k p(Z = k)p(\theta | Z = k)}{\int p(x | \theta) \sum_{k'} p(Z = k')p(\theta | Z = k')d\theta}
\]

\[
= \frac{\sum_k p(Z = k)p(x, \theta | Z = k)}{\sum_{k'} p(Z = k') \int p(x, \theta | Z = k')d\theta}
\]

\[
= \frac{\sum_k p(Z = k)p(\theta | x, Z = k)p(x | Z = k)}{\sum_{k'} p(Z = k')p(x | Z = k')}
\]

\[
= \sum_k \left[ \frac{p(Z = k)p(x | Z = k)}{\sum_{k'} p(Z = k')p(x | Z = k')} \right] p(\theta | x, Z = k)
\]

\[
= \sum_k p(Z = k | x) p(\theta | x, Z = k)
\]

Posterior mixing weights

Marginal likelihood

\[
p(x | Z = k) = \int p(x | \theta, Z = k)p(\theta | Z = k)d\theta
\]
Sequence logos

1. gtatcaccgcaccagtggtat 18.7
2. ataccactggcgggtgatac 18.7
3. tcaacaccgccagagataaa 19.1
4. ttatctcttgccggtgttga 19.1
5. ttatcaccgcagatgttta 16.7
6. taaccacatctcggtgtgataaa 16.7
7. ctatcaccgcagaggtataaa 18.4
8. ttatcctttgcgggtgatag 18.4
9. ctaacaccgtgcgtgttga 10.7
10. tcaacaccgcacgggtgtag 10.7
11. ttacctctgtgccggtgataaa 22.6
12. ttatcaccgcagagttta 22.6

12 Lambda cI and cro binding sites
Generative model

Inference goals:
Infer \( p(Z_t|X_{1:n,t}, \alpha, \pi) \)
and \( p(\theta_t|X_{1:n,t}, \alpha, \pi) \)

\[
\begin{align*}
  p(Z_t|\pi) &= \text{discrete}(\pi) \\
  p(\theta_t|Z_t = k, \alpha) &= \text{Dir}(\theta_t|\alpha_k) \\
  p(X_{it}|\theta_t) &= \text{discrete}(X_{it}|\theta_t) \\
  \alpha_1 &= (50, 1, 1, 1), \alpha_2 = (1, 50, 1, 1), \alpha_3 = (1, 1, 50, 1), \alpha_4 = (1, 1, 1, 50) \\
  \pi_k &= 1/4
\end{align*}
\]
Posterior

Sufficient statistics

\[ N_t = \left( \sum_{i=1}^{n} I(X_{it} = 1), \sum_{i=1}^{n} I(X_{it} = 2), \sum_{i=1}^{n} I(X_{it} = 3), \sum_{i=1}^{n} I(X_{it} = 4) \right) \]

Posterior on Z

\[
p(Z_t = k|D_t) = \frac{p(D_t|Z_t = k)p(Z_t = k)}{\sum_{k'} p(D_t|Z_t = k')p(Z_t = k')}
\]

\[
p(D_t|Z_t = k) = \frac{Z(N_t + \alpha_k)}{Z(\alpha_k)}
\]

\[
Z(\alpha) = \frac{\prod_j \Gamma(\alpha_j)}{\Gamma(\sum_j \alpha_j)}
\]

Posterior on \( \theta \)

\[
p(\theta_t | D_t) = \sum_{k=1}^{4} p(Z_t = k|D_t) \text{Dir}(\theta_t | \alpha_k + N_t)
\]

\[
E[\theta_t | D_t] = \sum_k E[\theta_t | D_t, Z_t = k] p(Z_t = k|D_t)
\]
Results

\[ E[\theta | \text{mix}] \]

\[ E[\theta | \text{unif}] \]
Which locations are conserved?
Generative model

\[ p(C_t) = \text{discrete}(C_t|p_c, 1 - p_c) \]

\[ p(Z_t|C_t = 1) = \text{discrete}(Z_t|\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0) \]

\[ p(Z_t|C_t = 0) = \text{discrete}(Z_t|0, 0, 0, 0, 1) \]

\[ p(\theta_t|Z_t = k) = \text{Dir}(\theta|\alpha_k) \]

\[ p(X_{it}|\theta_t) = \text{discrete}(X_{it}|\theta_t) \]

\[ \alpha_5 = (1, 1, 1, 1) \]

, \( \alpha_1, \ldots, \alpha_4 = 1 \)
Results

\[ p(C_t = 1|D_t) = \frac{p(C_t = 1)p(D_t|C_t = 1)}{\sum_{c=0}^{1} p(C_t = c)p(D_t|C_t = c)} \]

\[ p(D_t|C_t = c) = \sum_{k=1}^{5} p(D_t|Z_t = k)p(Z_t = k|C_t = c) \]