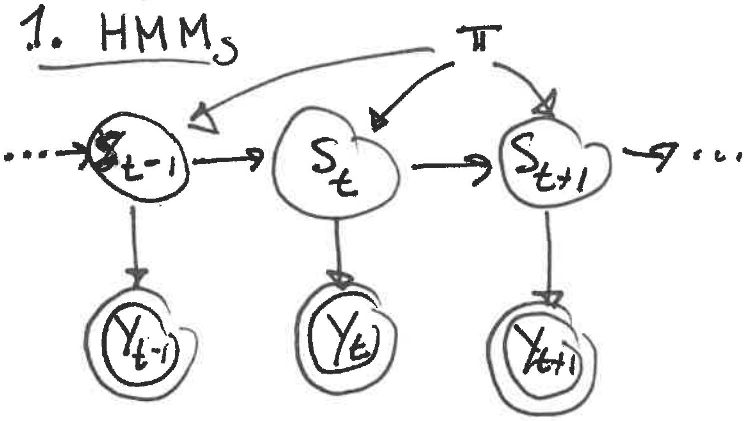
INFINITE HIDDEN MARKOV MODEL

1. HMMs



- States $s_t = 1, \dots, K$
- Transition:

  $s_t | s_{t-1} \sim \text{Cat}(\pi)$, $\pi_k \in [0,1]$, $\sum_k \pi_k = 1$

- Observation/emission:

  $y_t | s_t \sim p(\cdot | s_t)$

  Could be discrete/continuous

PARAMETER LEARNING

- Maximum likelihood

$$\max_{\pi} p(y_1, \dots, y_T | \pi) = \max_{\pi} \underbrace{\sum_{s_1} \dots \sum_{s_T}}_{\text{Sum over } T^K \text{ possible config.}} p_{\pi}(s_t | s_{t-1}) p(y_t | s_t)$$

- Baum-Welch alg. = Expectation Maximization for HMM

  ⇒ Uses Forward-Backward alg. (sum-product) in E-step to Calculate $p(S | y, \pi^{old})$

- Issues:

  ML estimate $\hat{\pi}$ tends to overfit training data

  ~~b) Need to specify number of states K~~

BAYESIAN HMM:

WRITE GENERAL FORM OF DIRICHLET

- Model transition prob. as random:

$$\pi \sim \text{Dir}(\beta/K, \dots \beta/K) = \frac{\Gamma(\beta)}{\Gamma(\beta/K)^K} \prod_{j=1}^{K} \pi_j^{\beta/K - 1}$$

~~Dirichlet / Multinomial conjugacy~~:

~~Let $\{c_1, \dots, c_n\}$ be discrete multinomial~~

- Dirichlet / Categorical conjugacy:

Let $\{c_i\}_{i=1}^{n} \overset{iid}{\sim} \mathrm{Cat}(\pi)$

<u>Joint</u>: $p(c_1, \dots, c_n | \pi) = \prod_{j=1}^{K} \pi_j^{n_j}$ with $n_j = \sum_{i=1}^{n} \mathbb{I}(c_i = j)$

<u>POSTERIOR</u>:

$$p(\pi | c_1^n, \beta) \propto p(\pi | \beta)\, p(c_1^n | \pi)$$

$$= \mathrm{DIR}(\pi | \beta) \prod_{i=1}^{n} \mathrm{Cat}(c_i | \pi)$$

$$= \frac{\Gamma(\beta)}{\Gamma(\beta/K)^K} \prod_{j=1}^{K} \pi_j^{\beta/K - 1} \prod_{j=1}^{K} \pi_j^{n_j}$$

$$= \frac{\Gamma(\beta)}{\Gamma(\beta/K)^K} \prod_{j=1}^{K} \pi_j^{(\beta/K + n_j - 1)} \quad \cancel{\propto \mathrm{DIR}(\pi | \beta/K + n_1, \dots, \beta/K + n_K)}$$

"Pseudo-counts"
Suff. Stats.

~~MARGINAL LIKELIHOOD~~

$$\cancel{p(c_1^n | \beta) = \int p(c_1^n | \pi)\, p(\pi | \beta)}\, d\pi$$

$$= \frac{\Gamma(\beta)}{\Gamma(\beta/K)^K} \frac{\Gamma(\sum_j \beta/K + n_j)}{\prod_j \Gamma(\beta/K + n_j)} \frac{\prod_j \Gamma(\beta/K + n_j)}{\Gamma(\sum_j \beta/K + n_j)} \prod_j \pi_j^{(\beta/K + n_j - 1)}$$

$$= \underbrace{\frac{\Gamma(\beta)}{\Gamma(\beta/K)^K} \frac{\prod_j \Gamma(\beta/K + n_j)}{\Gamma(\sum_j \beta/K + n_j)}}_{p(c|\beta)} \mathrm{DIR}(\pi | \beta/K + n_1, \dots, \beta/K + n_K)$$

pseudocounts
suff. stats.

2

MULTINOMIAL CONDITIONAL

$$p(c_d=j \mid c_{-d}) \propto p(c_d=j, c_{-d}) = \frac{\Gamma(\beta)}{\Gamma(n+\beta)} \prod_{j=1}^{k} \frac{\Gamma(n_j+\beta/k)}{\Gamma(\beta/k)}$$

$$= \frac{\Gamma(\beta)}{\Gamma(n+\beta)} \frac{\Gamma(n_{j,-d}+1+\beta/k)}{\Gamma(\beta/k)} \prod_{i\neq j} \frac{\Gamma(n_i+\beta/k)}{\Gamma(\beta/k)}$$

$$\neq \cancel{\frac{\Gamma(\beta)}{\Gamma(n+\beta)} \Gamma(n_{j,-d})\, \Gamma(n_{j,-d}}$$

$$= \frac{\Gamma(\beta)}{\Gamma(n+\beta)} \frac{\Gamma(n_{j,-d}+\beta/k)(n_{j,-d}+\beta/k)}{\Gamma(\beta/k)} \prod_{i\neq j} \frac{\Gamma(n_i+\beta/k)}{\Gamma(\beta/k)}$$

$$= (n_{j,-d}+\beta/k) \frac{\Gamma(\beta)}{\Gamma(n+\beta)} \prod_{i=1}^{k} \frac{\Gamma(n_{i,-d}+\beta/k)}{\Gamma(\beta/k)}$$

$$= (n_{j,-d}+\beta/k)\, \underbrace{p(c_{-d})}_{\text{const.}}$$

$$\cancel{p(c_d=j \mid c_{-d}) = n_{j,-d}+\beta/k}$$

$$\sum_{j=1}^{k} (n_{j,-d}+\beta/k) = \beta + n_{-d} = \beta + n - 1$$

$$p(c|\beta) = \frac{\Gamma(\beta)}{\Gamma(\beta+n)} \prod_{j=1}^{k} \frac{\Gamma(\beta/k + n_j)}{\Gamma(\beta/k)} = \int p(\pi|\beta)\, p(c_i^n|\pi)\, d\pi$$

=> We can analytically marginalize ~~in~~ π!

=> Only need pseudocounts (e.g. how many times j^{th} class appears)

~~Note~~

CATEGORICAL CONDITIONAL

=> SEPERATE SHEET

(2) DIRICHLET PROCESS: CHINESE RESTAURANT CONSTRUCTION

=> Let $K = \infty$ possible states

=> Let c_{-d} represent $\bar{K}$ states (finite)

=> Conditional dist'n:

pseudocounts

$$p(c_d = j \mid c_{-d}, \beta) = \begin{cases} \dfrac{n_{-d,j}}{n-1+\beta} & \text{if } j \in \{1, \dots, \bar{K}\} \\ \dfrac{\beta}{n-1+\beta} & \text{o.w. (new state)} \end{cases}$$

prior concentration

=> Verify integrates to 1:

$$\sum_{j=1}^{\infty} p(c_d = j \mid c_{-d}, \beta) = \left(\sum_{j=1}^{\bar{k}} \frac{n_{-d,j}}{n-1+\beta} \right) + \left(\sum_{j=\bar{k}+1}^{\infty} \frac{\beta}{n-1+\beta} \right)$$

$$= \frac{n-1}{n-1+\beta} + \frac{\beta}{n-1+\beta} = 1$$

=> Exchangeable!

=> ~~Leads~~ Leads to natural Gibbs sampling alg.

LOTS OF DIFF. WAYS TO REPRESENT/CONSTRUCT DPs

$n_{-d,1} = 4$ $\quad n_{-d,2} = 3$ $\quad n_{-d,3} = 1$ $\quad \dots$ $\quad \beta$

3) INFINITE HMM

⇒ Transition prob. π infinite matrix w/ DP prior

⇒ Marginalize π (prev. section) – work w/ Conditional instead

$$p(s_{t+1}=j \mid s_{-t+1}) = p(s_{t+1}=j \mid s_t, \underbrace{n}_{\text{Previous transition counts}})$$

⇒ Bi-level DP:

LEVEL 1:

$$p(s_{t+1}=j \mid s_t=i, \underset{\text{Concentration}}{n, \beta}, \underset{\text{self-transition Bias}}{\alpha}) = \begin{cases} \dfrac{n_{ii}+\alpha}{\sum_{j'} n_{ij'}+\beta+\alpha} & \text{if } i=j \\ \dfrac{n_{ij}}{\sum_{j'} n_{ij'}+\beta+\alpha} & \text{if } j\in\{1,\dots,\bar{K}\} \\ \dfrac{\beta}{\sum_{j'} n_{ij'}+\beta+\alpha} & \text{o.w. (new transition)} \end{cases}$$

LEVEL 2:

⇒ New transition to existing state $\{1,\dots,\bar{K}\}$

⇒ " " " new state

⇒ "Oracle" n_j^o Counts number of new transitions to $j\in\{1,\dots,\bar{K}\}$

$$p(s_{t+1}=j \mid s_t=i, n^o, \gamma) = \begin{cases} \dfrac{n_j^o}{\sum_{j'} n_{j'}^o+\gamma} & \text{if } j\in\{1,\dots,\bar{K}\} \\ \dfrac{\gamma}{\sum_{j'} n_{j'}^o+\gamma} & \text{o.w. (new trans to new state)} \end{cases}$$

⇒ Prob transition to new state $\propto \beta\gamma$

⇒ 3 HYPERPARAMETERS: $\{\alpha, \beta, \gamma\}$