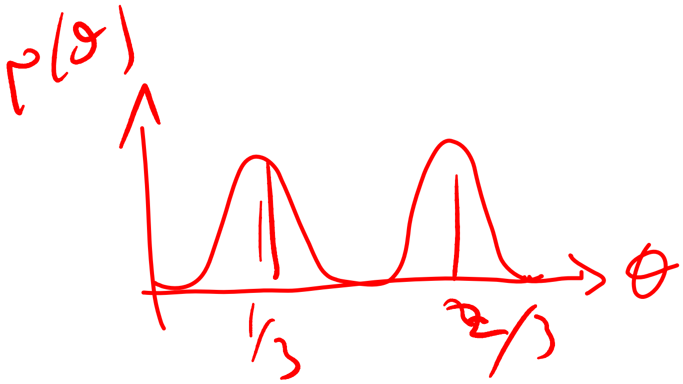
CS340 Machine learning
Mixture priors

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,…,K} is a latent variable which specifies which prior mixture component to use

$$\begin{tikzcd} & \pi \arrow[d] \\ & Z \arrow[d] \\ \alpha \arrow[r] & \theta \arrow[d] \\ & x \end{tikzcd}$$

$$\begin{aligned} p(Z = k|\boldsymbol{\pi}) &= \pi_k \\ p(\theta|Z = k, \boldsymbol{\alpha}) &= p(\theta|\alpha_k) \\ p(x|\theta) &= p(x|\theta) \end{aligned}$$

Example: Beta(θ|10, 20) or Beta(θ | 20, 10)

Posterior is also a mixture

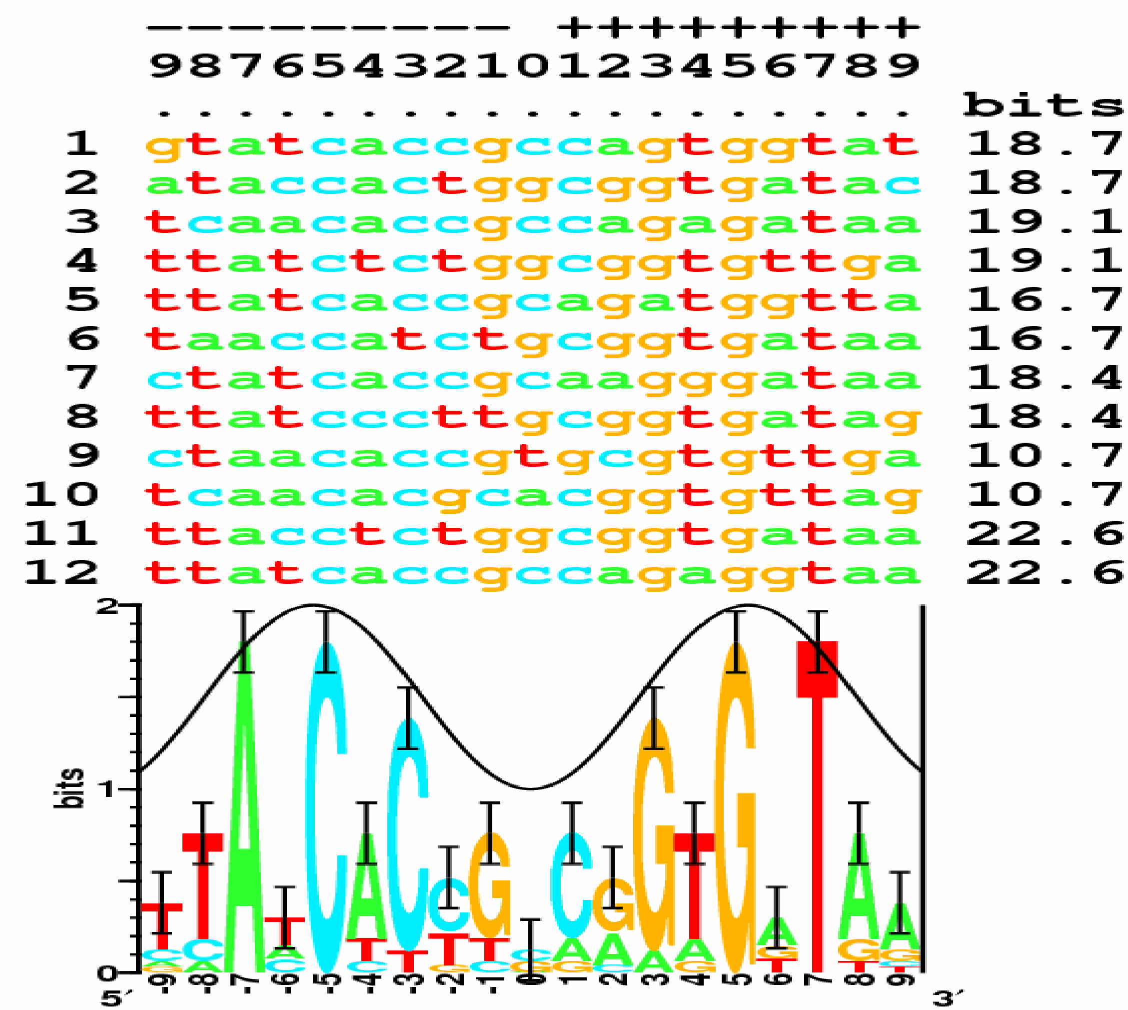
$$
\begin{aligned}
p(\theta|x) &= \frac{p(x|\theta)p(\theta)}{p(x)} \\
&= \frac{p(x|\theta)\sum_k p(Z=k)p(\theta|Z=k)}{\int p(x|\theta)\sum_{k'} p(Z=k')p(\theta|Z=k')d\theta} \\
&= \frac{\sum_k p(Z=k)p(x,\theta|Z=k)}{\sum_{k'} p(Z=k')\int p(x,\theta|Z=k')d\theta} \\
&= \frac{\sum_k p(Z=k)p(\theta|x,Z=k)p(x|Z=k)}{\sum_{k'} p(Z=k')p(x|Z=k')} \\
&= \sum_k \left[\frac{p(Z=k)p(x|Z=k)}{\sum_{k'} p(Z=k')p(x|Z=k')}\right] p(\theta|x,Z=k) \\
&= \sum_k p(Z=k|x)p(\theta|x,Z=k)
\end{aligned}
$$

posterior mixing weights

Marginal likelihood

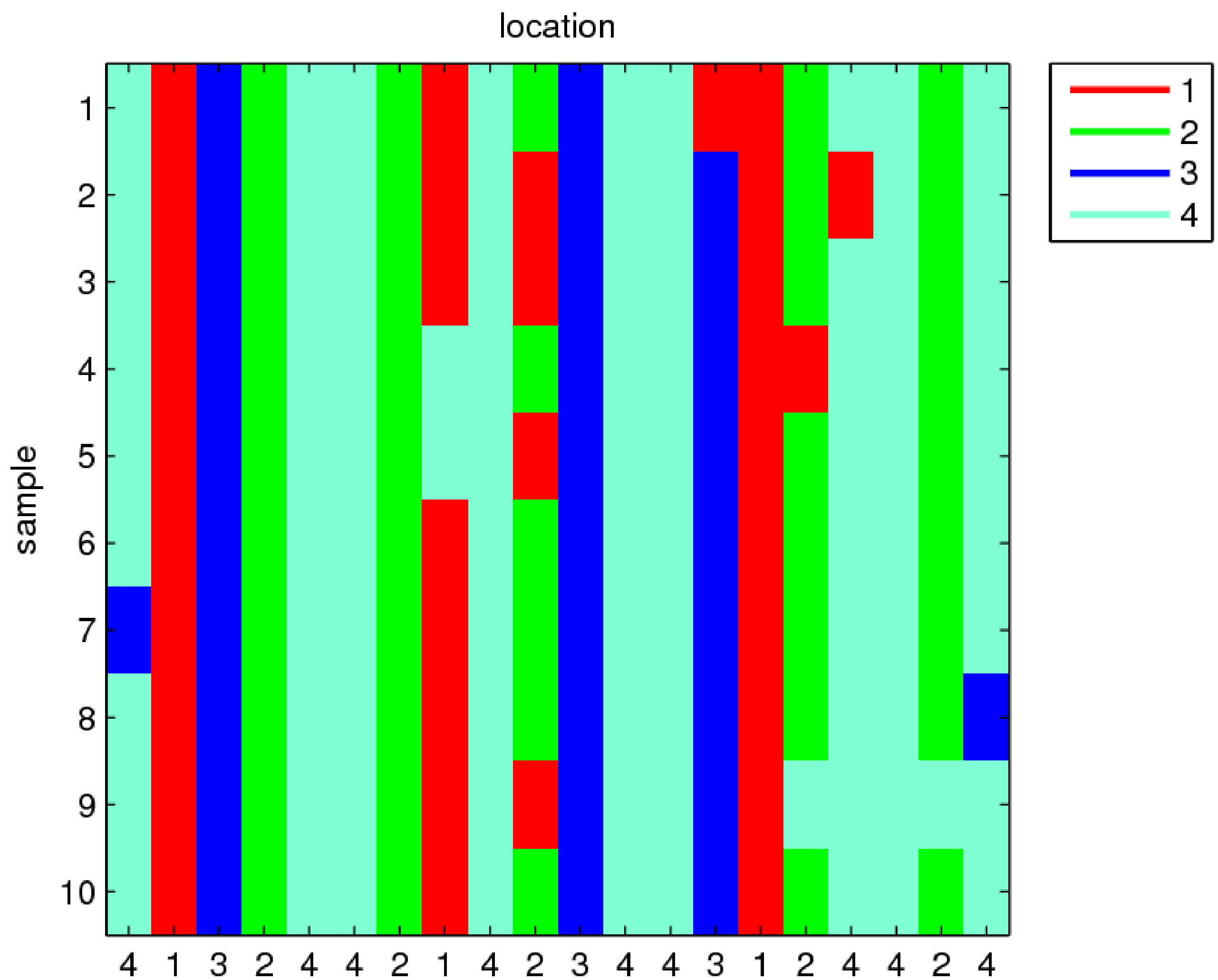
$$p(x|Z=k) = \int p(x|\theta,Z=k)p(\theta|Z=k)d\theta$$

Sequence logos

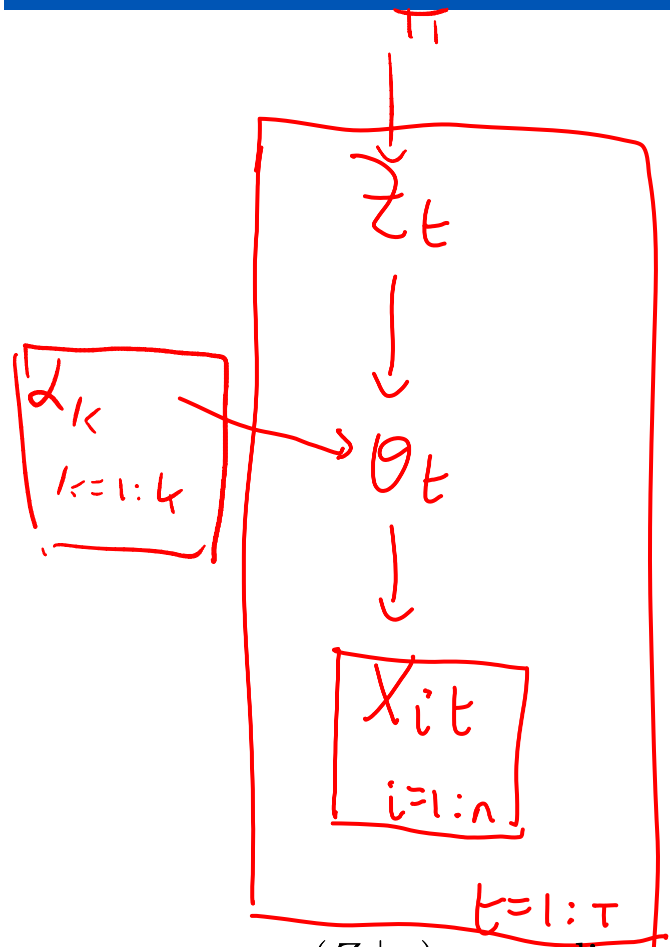


12 Lambda cI and cro binding sites

Data



Generative model



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

$$
\begin{aligned}
p(Z_t|\boldsymbol{\pi}) &= \text{discrete}(\boldsymbol{\pi}) \\
p(\boldsymbol{\theta}_t|Z_t = k, \boldsymbol{\alpha}) &= Dir(\boldsymbol{\theta}_t|\boldsymbol{\alpha}_k) \\
p(X_{it}|\boldsymbol{\theta}_t) &= \text{discrete}(X_{it}|\boldsymbol{\theta}_t) \\
\boldsymbol{\alpha}_1 &= (50,1,1,1), \boldsymbol{\alpha}_2 = (1,50,1,1), \boldsymbol{\alpha}_3 = (1,1,50,1), \boldsymbol{\alpha}_4 = (1,1,1,50) \\
\pi_k &= 1/4
\end{aligned}
$$

Posterior

Sufficient statistics

$$\mathbf{N}_t = \left(\sum_{i=1}^{n} I(X_{it} = 1), \sum_{i=1}^{n} I(X_{it} = 2), \sum_{i=1}^{n} I(X_{it} = 3), \sum_{i=1}^{n} I(X_{it} = 4) \right)$$

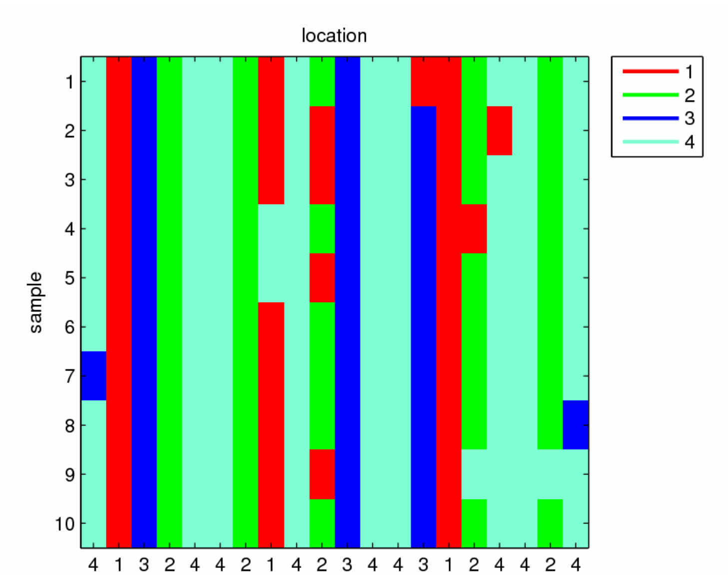
Posterior on Z

$$\begin{aligned} p(Z_t = k | D_t) &= \frac{p(D_t | Z_t = k) p(Z_t = k)}{\sum_{k'} p(D_t | Z_t = k') p(Z_t = k')} \\ p(D_t | Z_t = k) &= \frac{Z(\mathbf{N}_t + \boldsymbol{\alpha}_k)}{Z(\boldsymbol{\alpha}_k)} \\ Z(\boldsymbol{\alpha}) &= \frac{\prod_j \Gamma(\alpha_j)}{\Gamma(\sum_j \alpha_j)} \end{aligned}$$

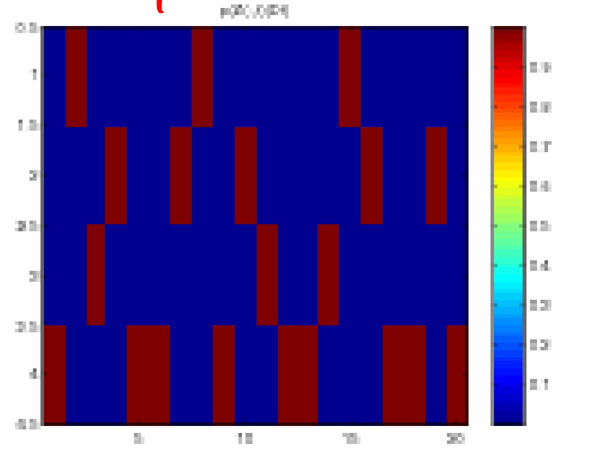
Posterior on θ

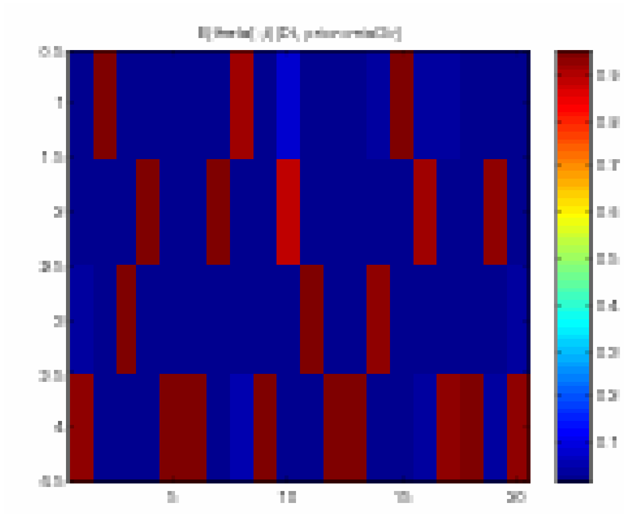
$$\begin{aligned} p(\boldsymbol{\theta}_t | D_t) &= \sum_{k=1}^{4} p(Z_t = k | D_t) Dir(\boldsymbol{\theta}_t | \boldsymbol{\alpha}_k + \mathbf{N}_t) \\ E[\boldsymbol{\theta}_t | D_t] &= \sum_{k} E[\boldsymbol{\theta}_t | D_t, Z_t = k] p(Z_t = k | D_t) \end{aligned}$$

Results

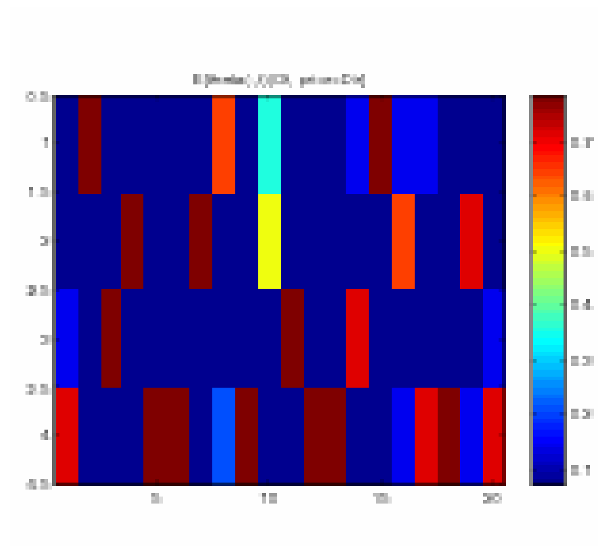


$p(z)$



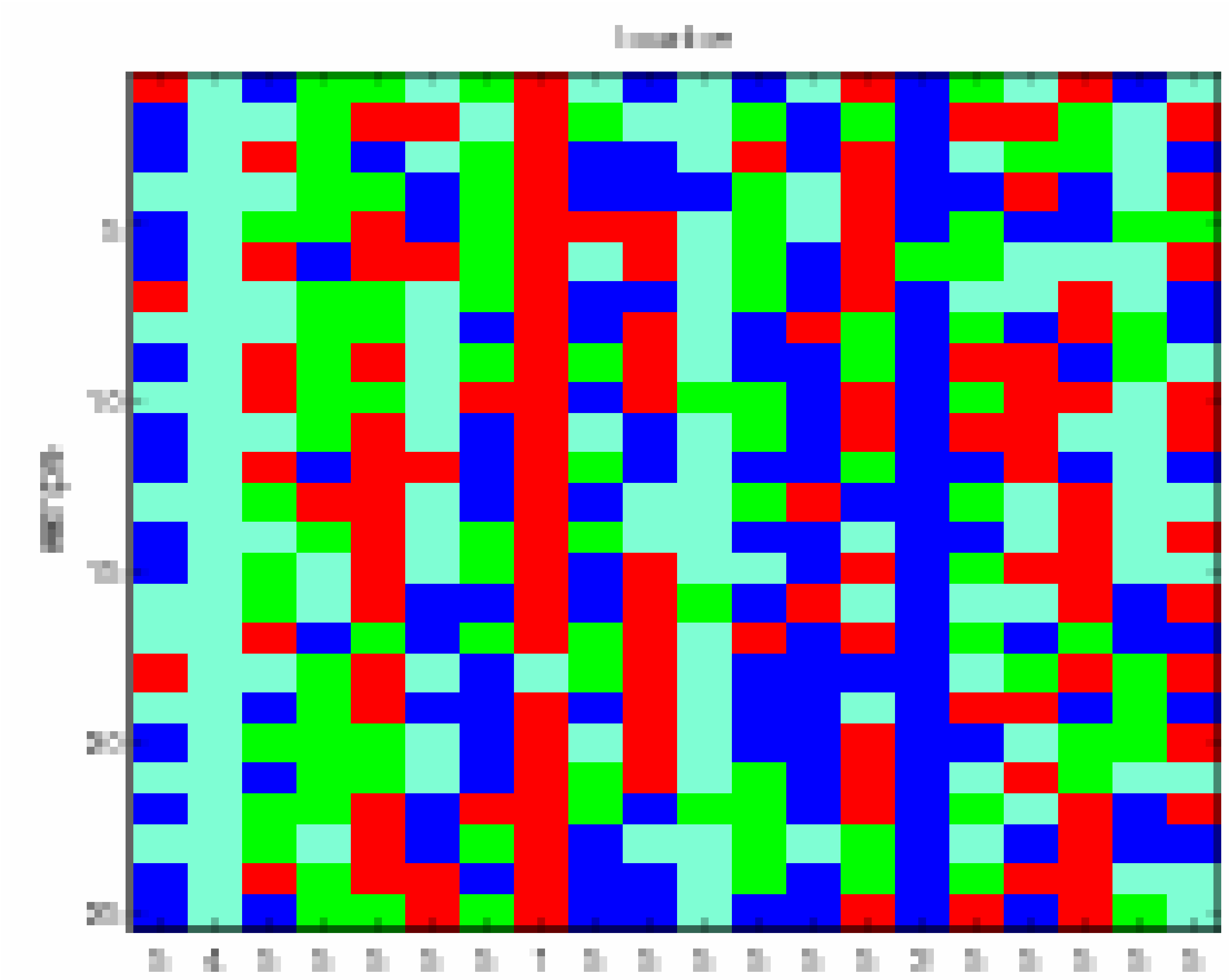


$E[\theta|mix]$

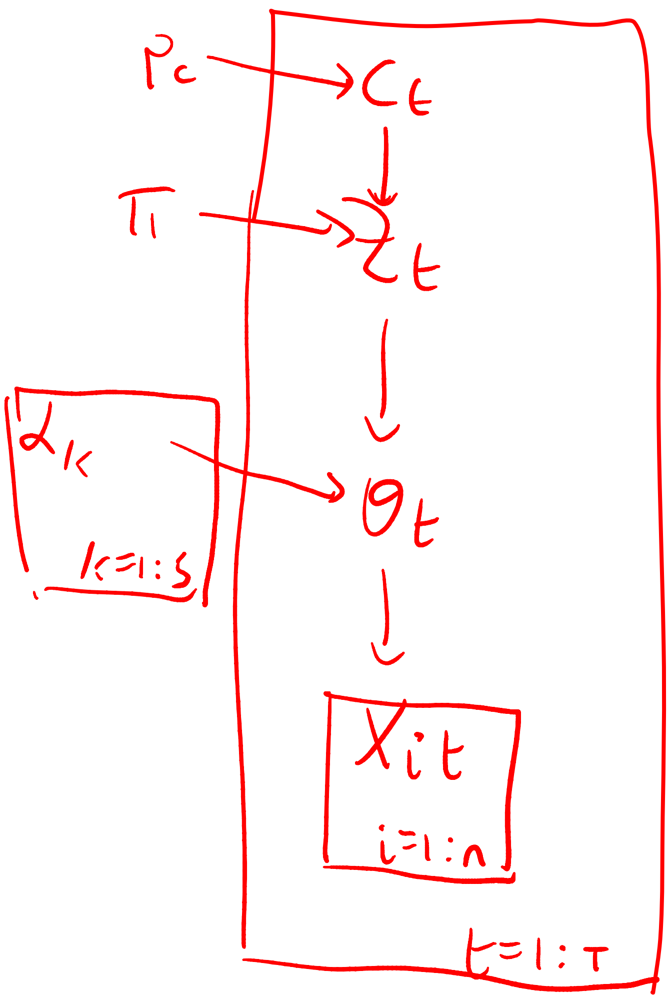


$E[\theta|unif]$

Which locations are conserved?

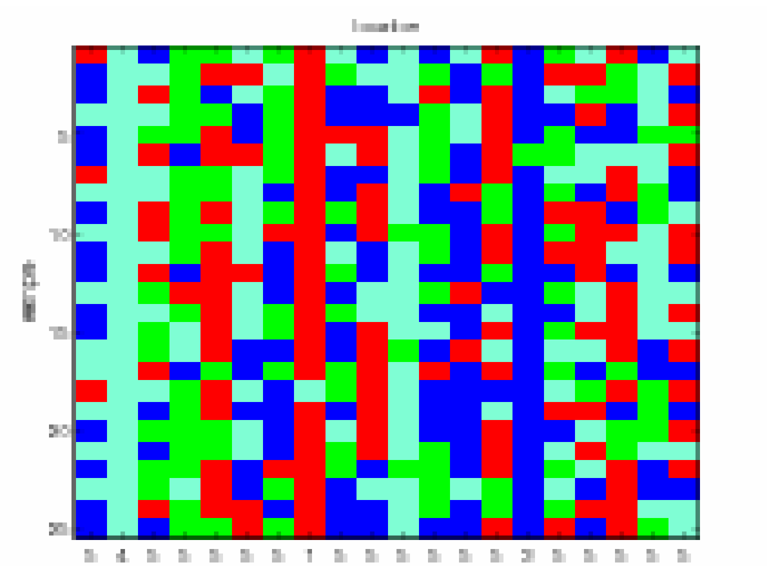


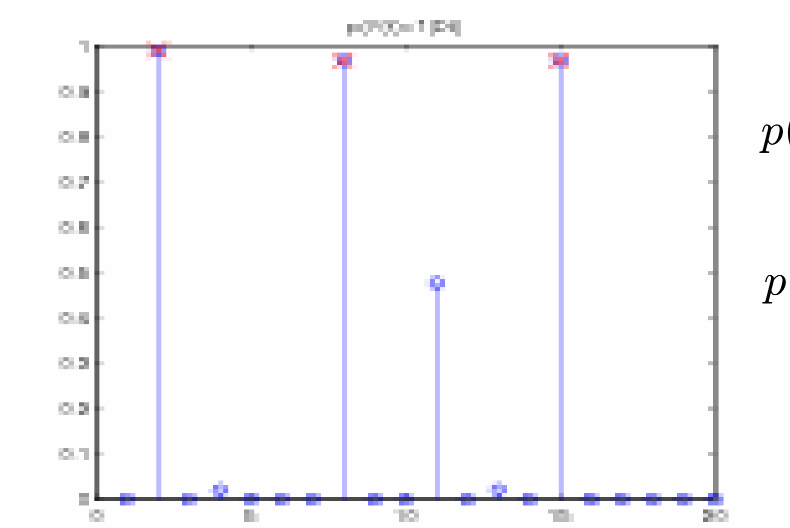
Generative model



$$
\begin{aligned}
p(C_t) &= \text{discrete}(C_t|(p_c, 1-p_c)) \\
p(Z_t|C_t=1) &= \text{discrete}(Z_t|(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0)) \\
p(Z_t|C_t=0) &= \text{discrete}(Z_t|(0,0,0,0,1)) \\
p(\boldsymbol{\theta}_t|Z_t=k) &= Dir(\boldsymbol{\theta}|\boldsymbol{\alpha}_k) \\
p(X_{it}|\boldsymbol{\theta}_t) &= \text{discrete}(X_{it}|\boldsymbol{\theta}_t) \\
\boldsymbol{\alpha}_5 = (1,1,1,1) & \quad , \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_4 = \mathbf{1}
\end{aligned}
$$

Results





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

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