

why it could be defined as below:

 $\boldsymbol{z}_n = [0,0,\ldots,1,\ldots,0]^T$, represent probability of length = K states, only when $z_{nk} = 1$, it would be counted into p, **random variable**

$$egin{aligned} p\left(oldsymbol{z_1}|oldsymbol{\pi}
ight) &= \prod_{k=1}^K \pi_k^{z_{1k}} \ p\left(oldsymbol{z_n}|oldsymbol{z_{n-1}},oldsymbol{A}
ight) &= \prod_{j=1}^K \prod_{k=1}^K A_{j,k}^{z_{n-1,j}z_{nk}} \ p\left(oldsymbol{x_n}|oldsymbol{z_n},oldsymbol{\phi}
ight) &= \prod_{j=1}^K \prod_{t=1}^{any\,dimension} \phi_j(t)^{z_{nj}x_n(t)} \end{aligned}$$

Define \vec{p} constant matrix

$$ec{p}(oldsymbol{z}_{n}|oldsymbol{ heta}) \equiv egin{bmatrix} p(oldsymbol{z}_{n}|oldsymbol{ heta}) oldsymbol{z}_{n} = \operatorname{state} 1 \ dots \ p(oldsymbol{z}_{n}|oldsymbol{ heta}) oldsymbol{z}_{n} = \operatorname{state} K \end{bmatrix} = egin{bmatrix} \sum_{oldsymbol{z}_{n}} p\left(oldsymbol{z}_{n}|oldsymbol{X}, oldsymbol{ heta} & \operatorname{old} \right) oldsymbol{z}_{n} \end{bmatrix} \quad \operatorname{state} k : oldsymbol{z}_{n} = egin{bmatrix} 0 \ dots \ p(oldsymbol{z}_{n}|oldsymbol{z}_{n} = \operatorname{state} k \end{bmatrix} = oldsymbol{\pi} \\ ec{p}\left(oldsymbol{z}_{n}|oldsymbol{z}_{n-1}, oldsymbol{ heta}
ight) igg| oldsymbol{z}_{n} = \operatorname{state} i, oldsymbol{z}_{n-1} = \operatorname{state} j \end{bmatrix} = oldsymbol{A} \\ ec{p}\left(oldsymbol{x}_{n}|oldsymbol{z}_{n}, oldsymbol{ heta}
ight) igg| oldsymbol{x}_{n} = \operatorname{state} i, oldsymbol{z}_{n} = \operatorname{state} j \end{bmatrix} = oldsymbol{\phi} \end{aligned}$$

where, for short: write $\vec{p}({m z_n}) \equiv \vec{p}({m z_n}|{m heta}), \vec{p}({m x_n}) \equiv \vec{p}({m x_n}|{m heta})$

$$egin{aligned} ec{p}(oldsymbol{z_1}) &= oldsymbol{\pi} \ ec{p}(oldsymbol{z_n}) &= oldsymbol{A} ec{p}(oldsymbol{z_{n-1}}) \ ec{p}(oldsymbol{x_n}) &= oldsymbol{\phi} \cdot ec{p}(oldsymbol{z_n}) \end{aligned}$$

 $\operatorname{size}(p(oldsymbol{x_n}|oldsymbol{\phi}))$ = $any\ dimension imes K$, $oldsymbol{\phi} = [\phi_1,\phi_2,\ldots,\phi_j,\ldots\phi_K]$

 ϕ_k is probability distribution sequence of k th state, $any\ dimension imes 1$

 $oldsymbol{x}_n$ is probability distribution sequence, $any\ dimension imes 1$

 $oldsymbol{z_n}$ is probability distribution sequence, K imes 1

In P(z,theta), z is observation value;

In $z_n = A z_{n-1}$, z is random variable;

Thus

$$p\left(oldsymbol{x_n} | oldsymbol{z_n}, oldsymbol{\phi}
ight) = \prod_{j=1}^K \prod_{t=1}^{any \ dimension} \phi_j(t)^{z_{nj}x_n(t)}$$

Goal

$$egin{aligned} p(x_1,\dots,x_n,z_1,\dots,z_n|m{\pi},m{A},m{\phi}) &= p(x_1,\dots,x_n,z_1,\dots,z_n|m{ heta}) \ &= p(x_n|x_1,\dots,x_{n-1},z_1,\dots,z_n,m{ heta}) \cdot p(x_1,\dots,x_{n-1},z_1,\dots,z_n|m{ heta}) \ &= [p(x_n|z_n,m{\phi})] \cdot p(x_1,\dots,x_{n-1},z_1,\dots,z_n|m{ heta}) \ &= [p(x_n|z_n,m{\phi})] \cdot p(z_n|x_1,\dots,x_{n-1},z_1,\dots,z_{n-1},m{ heta}) \cdot p(x_1,\dots,x_{n-1},z_1,\dots,z_{n-1}|m{ heta}) \ &= [p(x_n|z_n,m{\phi})] \cdot [p(z_n|z_{n-1},m{A})] \cdot p(x_1,\dots,x_{n-1},z_1,\dots,z_{n-1}|m{ heta}) \end{aligned}$$

thus we get

$$p(x_1,\ldots,x_n,z_1,\ldots,z_n|m{ heta}) = [p(x_n|z_n,m{\phi})]\cdot[p(z_n|z_{n-1},m{A})]\cdot p(x_1,\ldots,x_{n-1},z_1,\ldots,z_{n-1}|m{ heta})$$

$$p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta}) \equiv p(\boldsymbol{x_1},\ldots,\boldsymbol{x_N},\boldsymbol{z_1},\ldots,\boldsymbol{z_N}|\boldsymbol{\theta}) = [\prod_{n=1}^N p(\boldsymbol{x_n}|\boldsymbol{z_n},\boldsymbol{\phi})] \cdot [\prod_{n=2}^N p(\boldsymbol{z_n}|\boldsymbol{z_{n-1}},\boldsymbol{A})] \cdot p\left(\boldsymbol{z_1}|\boldsymbol{\pi}\right)$$

Here

$$egin{aligned} p\left(oldsymbol{x_n}|oldsymbol{z_n},oldsymbol{\phi}
ight) &= \prod_{j=1}^K \prod_{t=1}^{any\,dimension} \phi_j(t)^{z_{nj}x_n(t)} \ p\left(oldsymbol{z_n}|oldsymbol{z_{n-1}},oldsymbol{A}
ight) &= \prod_{j=1}^K \prod_{k=1}^K A_{k,j}^{z_{n-1,j}z_{nk}} \ p\left(oldsymbol{z_1}|oldsymbol{\pi}
ight) &= \prod_{k=1}^K \pi_k^{z_{1k}} \end{aligned}$$

So

$$egin{aligned} & \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) \ &= \sum_{n=1}^{N} \sum_{j=1}^{K} \sum_{t=1}^{any\,dimension} z_{nj} x_n(t) \ln \phi_j(t) + \sum_{n=2}^{N} \sum_{k=1}^{K} \sum_{j=1}^{K} z_{n-1,j} z_{n,k} \ln A_{kj} + \sum_{k=1}^{K} z_{1k} \ln \pi_k \ &= \sum_{n=1}^{N} oldsymbol{z}_n^T < \ln oldsymbol{\phi}, oldsymbol{x}_n > + \sum_{n=2}^{N} oldsymbol{z}_{n-1}^T \ln oldsymbol{A}^T \cdot oldsymbol{z}_n + \ln oldsymbol{\pi}^T \cdot oldsymbol{z}_1 \end{aligned}$$

if t is continuous

$$egin{aligned} < \ln oldsymbol{\phi}, oldsymbol{x_n}> &= \int_{-\infty}^{+\infty} \ln oldsymbol{\phi}(oldsymbol{t}) \cdot x_n(t) dt \ oldsymbol{\phi} &= egin{bmatrix} \phi_1(t) \ \phi_2(t) \ dots \ \phi_K(t) \end{bmatrix} \end{aligned}$$

if m is discrete

$$\langle \ln \phi, oldsymbol{x_n}
angle = \ln oldsymbol{\phi}^T \cdot oldsymbol{x_n} \ oldsymbol{\phi} = egin{bmatrix} \phi_{11} & \cdots & \phi_{1m} & \cdots \phi_{1M} \ \phi_{21} & \cdots & \phi_{2m} & \cdots \phi_{2M} \ dots & & dots \ \phi_{K1} & \cdots & \phi_{Km} & \cdots \phi_{KM} \end{bmatrix}^T M imes K ext{matrix} \ oldsymbol{x_n} : M imes 1 ext{ matrix}$$

Only $\boldsymbol{x_n}$ is known, $n \in \{1, \dots, N\}$

then

$$egin{aligned} P(Y=y)dy &\equiv P(X=x)dx = p(x)dx, \quad y=f(x) \ &\Longrightarrow P(Y=y) = p(x)rac{dx}{df(x)} \ &\Longrightarrow E[Y] = \int_{-\infty}^{+\infty} y\cdot P(Y=y)dy = \int_{-\infty}^{+\infty} f(x)p(x)dx \end{aligned}$$

期望最大化算法,或者EM算法

$$\max \quad \ln p(\boldsymbol{X}|\boldsymbol{\theta}) = \mathcal{L}(q, \boldsymbol{\theta}) + \mathrm{KL}(q||p)$$

where

$$\begin{split} \mathcal{L}(q, \boldsymbol{\theta}) &= \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln \left\{ \frac{p(\boldsymbol{X}, \boldsymbol{Z}|\boldsymbol{\theta})}{q(\boldsymbol{Z})} \right\} \\ \text{KL}(q \| p) &= -\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \ln \left\{ \frac{p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{\theta})}{q(\boldsymbol{Z})} \right\} \end{split}$$

Where

- $q(\boldsymbol{Z})$ is \forall arbitrary distribution for \boldsymbol{Z}
- $\mathcal{L}(q, \boldsymbol{\theta})$ 概率分布 $q(\boldsymbol{Z})$ 的一个泛函, $\boldsymbol{\theta}$ 函数
- KL(q||p) 概率分布 $q(\mathbf{Z})$ 的一个泛函, $\boldsymbol{\theta}$ 函数,KL散度 of $p(Z|X,\theta),\ q(Z),\geq 0$

So

$$egin{aligned} & \ln p(oldsymbol{X}|oldsymbol{ heta}) \geq \mathcal{L}(q,oldsymbol{ heta}) = \sum_{oldsymbol{Z}} q(oldsymbol{Z}) \ln iggl\{ rac{p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta}}{q(oldsymbol{Z})} iggr\}, \quad orall q(oldsymbol{Z}) \ & ext{if } q(oldsymbol{Z}) \Leftarrow p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}), ext{here } \operatorname{KL}(q||p) = 0 \ & ext{ln } p(oldsymbol{X}|oldsymbol{ heta}) = \mathcal{L}(q,oldsymbol{ heta}) = \sum_{oldsymbol{Z}} p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta}) \ln iggl\{ rac{p(oldsymbol{X}, oldsymbol{Z}|oldsymbol{ heta})}{p(oldsymbol{Z}|oldsymbol{X}, oldsymbol{ heta})} iggr\} \end{aligned}$$

notice: X is fixed, Z, θ are variables,转而用 $q(Z), \theta$ 替代;

调整法

$$\ln p(m{X}|m{ heta}) = \mathcal{L}(q,m{ heta}) + \mathrm{KL}(q|p)$$

1. fix θ , change q(Z) [E步骤]

$$egin{aligned} q(oldsymbol{Z}) &\Leftarrow p(oldsymbol{Z}|oldsymbol{X},oldsymbol{ heta}) \ \ln p(oldsymbol{X}|oldsymbol{ heta}) &= \mathcal{L}(q,oldsymbol{ heta}) + \mathrm{KL}(q|p) = \mathrm{const}$$
与q无关 $\mathrm{KL}(q|p)\downarrow \Leftarrow 0, \quad \mathrm{update} \ \mathcal{L}(q,oldsymbol{ heta}) \uparrow \ \ln p(oldsymbol{X}|oldsymbol{ heta}) &== \mathcal{L}(q,oldsymbol{ heta}) \ oldsymbol{ heta} \ oldsymbol{ heta}$

2. fix q(Z), change θ [M步骤]

ignore influence of
$$\theta$$
 in $\mathrm{KL}(q|p) = -\sum_{m{Z}} q(m{Z}) \ln \left\{ \frac{p(m{Z}|X, \theta)}{q(m{Z})} \right\}$
$$\mathcal{L}(q, m{\theta}) = \sum_{m{Z}} q(m{Z}) \ln \left\{ \frac{p(m{X}, m{Z}| m{\theta})}{q(m{Z})} \right\}$$

$$= \sum_{m{Z}} p(m{Z}|m{X}, m{\theta}^{\mathrm{old}}) \ln \left\{ \frac{p(m{X}, m{Z}| m{\theta})}{p(m{Z}|m{X}, m{\theta}^{\mathrm{old}})} \right\}$$

$$= \sum_{m{Z}} p\left(m{Z}|m{X}, m{\theta}^{\mathrm{old}}\right) \ln p(m{X}, m{Z}| m{\theta}) - \mathrm{const}$$

$$oldsymbol{Q}\left(oldsymbol{ heta}, oldsymbol{ heta}^{
m old}
ight) \equiv \sum_{oldsymbol{Z}} p\left(oldsymbol{Z} | oldsymbol{X}, oldsymbol{ heta}^{
m old}
ight) \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta})$$

get

$$oldsymbol{ heta} \Leftarrow rgmax \mathcal{L}(q, \overline{oldsymbol{ heta}}) \ = rgmax oldsymbol{Q} \left(\overline{oldsymbol{ heta}}, oldsymbol{ heta}^{
m old}
ight) \ ext{update } \mathcal{L}(q, oldsymbol{ heta}) \uparrow$$

update

$$egin{aligned} \operatorname{KL}(q|p) \uparrow = -\sum_{oldsymbol{Z}} q(oldsymbol{Z}) \ln \left\{ rac{p(oldsymbol{Z}|oldsymbol{X},oldsymbol{ heta})}{q(oldsymbol{Z})}
ight\} > 0 \ & \ln p(oldsymbol{X}|oldsymbol{ heta}) \uparrow = \mathcal{L}(q,oldsymbol{ heta}) \uparrow + \operatorname{KL}(q|p) \uparrow \end{aligned}$$

Conclusion

因为1,2中
$$\mathcal{L}(q, \boldsymbol{\theta}) \uparrow$$
 均上升;

且1中
$$\ln p(\boldsymbol{X}|\boldsymbol{\theta}) = \mathcal{L}(q,\boldsymbol{\theta}) \uparrow +0 \downarrow = \text{const}$$

$$2 \pitchfork \ln p(oldsymbol{X} | oldsymbol{ heta}) \uparrow = \mathcal{L}(q, oldsymbol{ heta}) \uparrow + \mathrm{KL}(q|p) \uparrow$$

• 所以在一个循环1,2中
$$\ln p(\pmb{X}|\pmb{\theta})$$
 ↑下界 $\mathcal{L}(q,\pmb{\theta})$ ↑

consider $Q(\theta^{old}, \theta)$

Use
$$\sum$$
 to delete variables, $\sum\limits_{z_1} p(z_1,z_2) = P(z_2)$

X is fix observation value;

Z is [probability distribution] random variable, $\ln p(X,Z|\theta)$ is random variable

 $\sum_{m{Z}} p\left(m{Z}|m{X}, m{ heta}^{
m old}
ight)$ is [probability distribution] random variable **of** Z [probability distribution] random variable

$$\begin{split} \boldsymbol{Q}\left(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{ old }}\right) &\equiv \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \ln p(\boldsymbol{X},\boldsymbol{Z}|\boldsymbol{\theta}) \\ &= \sum_{\boldsymbol{Z}} p\left(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \left\{ \sum_{n=1}^{N} \boldsymbol{z}_{n}^{T} < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > + \sum_{n=2}^{N} \boldsymbol{z}_{n-1}^{T} \ln \boldsymbol{A}^{T} \cdot \boldsymbol{z}_{n} + \ln \boldsymbol{\pi}^{T} \cdot \boldsymbol{z}_{1} \right\} \\ &= \left\{ \sum_{\boldsymbol{z}_{n}} p\left(\boldsymbol{z}_{n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \sum_{n=1}^{N} \boldsymbol{z}_{n}^{T} < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > \right\} \\ &+ \left\{ \sum_{\boldsymbol{z}_{n-1},\boldsymbol{z}_{n}} p\left(\boldsymbol{z}_{n-1},\boldsymbol{z}_{n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \sum_{n=2}^{N} \boldsymbol{z}_{n-1}^{T} \ln \boldsymbol{A}^{T} \cdot \boldsymbol{z}_{n} \right\} \\ &+ \left\{ \sum_{\boldsymbol{z}_{1}} p\left(\boldsymbol{z}_{1}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \boldsymbol{z}_{1}^{T} \ln \boldsymbol{\pi} \right\} \\ &= \left\{ \sum_{n=1}^{N} \left[\sum_{\boldsymbol{z}_{n}} p\left(\boldsymbol{z}_{n}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \boldsymbol{z}_{n}^{T} \right] < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > \right\} \\ &+ \left\{ \left[\sum_{\boldsymbol{z}_{1}} p\left(\boldsymbol{z}_{1}|\boldsymbol{X},\boldsymbol{\theta}^{\text{ old }}\right) \boldsymbol{z}_{1}^{T} \right] \ln \boldsymbol{\pi} \right\} \\ &= \sum_{n=1}^{N} \boldsymbol{\gamma}_{n}^{T} < \ln \boldsymbol{\phi}, \boldsymbol{x}_{n} > + \sum_{n=2}^{N} \operatorname{tr}\left(\boldsymbol{\xi}_{n}^{T} \ln \boldsymbol{A}\right) + \boldsymbol{\gamma}_{1}^{T} \ln \boldsymbol{\pi} \end{split}$$

where

$$egin{aligned} &\sum_{oldsymbol{z}_{n-1},oldsymbol{z}_n} p\left(oldsymbol{z}_{n-1},oldsymbol{z}_n|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) \sum_{n=2}^N oldsymbol{z}_{n-1}^T \ln oldsymbol{A}^T \cdot oldsymbol{z}_n \ &= \sum_{n=2}^N \operatorname{tr}\left(p\left(oldsymbol{z}_{n-1},oldsymbol{z}_n|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight) \left[oldsymbol{z}_{n-1}^T \ln oldsymbol{A}^T \cdot oldsymbol{z}_n
ight]^T
ight) \ &= \sum_{n=2}^N \operatorname{tr}\left(\left[oldsymbol{z}_{n-1}p\left(oldsymbol{z}_{n-1},oldsymbol{z}_n|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight)oldsymbol{z}_n^T
ight] \cdot \ln oldsymbol{A}
ight) \ &= \sum_{n=2}^N \sum_{n=2} \left\{\left[oldsymbol{z}_np\left(oldsymbol{z}_{n-1},oldsymbol{z}_n|oldsymbol{X},oldsymbol{ heta}^{
m old}
ight)oldsymbol{z}_{n-1}^T
ight] \odot \ln oldsymbol{A}
ight\} \end{aligned}$$

其中, 因为X, θ ^{old} 分别为已知、固定 在 M步骤

$$egin{aligned} m{\gamma_n} &\equiv \left[\sum_{m{z_n}} p\left(m{z_n} | m{X}, m{ heta}^{
m old}
ight) m{z_n}
ight]$$
 为常数,是期望 $E[m{z_n} | m{X}, m{ heta}^{
m old}
ight]$,大小 $K imes 1$ $m{\xi_n} &\equiv \left[m{z_n} p\left(m{z_{n-1}}, m{z_n} | m{X}, m{ heta}^{
m old}
ight) m{z_{n-1}}^T
ight]$ 为常数,是期望 $E[m{z_n} m{z_{n-1}}^T | m{X}, m{ heta}^{
m old}
ight]$,大小 $K imes K$

下面求解

$$egin{aligned} \max_{oldsymbol{ heta}} & oldsymbol{Q}\left(oldsymbol{ heta}, oldsymbol{ heta} ext{ old}
ight) = \sum_{oldsymbol{Z}} p\left(oldsymbol{Z} | oldsymbol{X}, oldsymbol{ heta} ext{ old}
ight) \ln p(oldsymbol{X}, oldsymbol{Z} | oldsymbol{ heta}) \\ &= \sum_{oldsymbol{n}=1}^N oldsymbol{\gamma}_{oldsymbol{n}}^T < \ln oldsymbol{\phi}, oldsymbol{x}_{oldsymbol{n}} > + \sum_{oldsymbol{n}=2}^N \operatorname{tr}\left(oldsymbol{\xi}_{oldsymbol{n}}^T \ln oldsymbol{A}
ight) + oldsymbol{\gamma}_{oldsymbol{1}}^T \ln oldsymbol{\pi} \end{split}$$

subject to

s.t.
$$\boldsymbol{\gamma}_n^T \cdot 1 = 1$$
 tr $(\boldsymbol{\xi}_n^T \cdot 1) = 1$
 $1 \cdot T \cdot \boldsymbol{\phi} = 1 \cdot T$ $1 \cdot T \cdot \boldsymbol{A} = 1 \cdot T$ $1 \cdot T \cdot \boldsymbol{\pi} = 1$

with Lagrange method:

$$egin{aligned} L &\equiv \left\{ \sum_{n=1}^{N} oldsymbol{\gamma_n^T} < \ln oldsymbol{\phi}, oldsymbol{x_n} > + \sum_{n=2}^{N} \operatorname{tr}\left(oldsymbol{\xi_n^T} \ln oldsymbol{A}
ight) + oldsymbol{\gamma_1^T} \ln oldsymbol{\pi}
ight\} \ &- \sum_{t=1}^{any \ dimension} u_t \left\{ 1.^T \cdot < oldsymbol{\phi}, \delta(t) > -1
ight\} \ &- \sum_{k=1}^{K} v_k \left\{ 1.^T oldsymbol{A} \cdot \sigma_k - 1
ight\} \ &- w_1 \left\{ 1.^T oldsymbol{\pi} - 1
ight\} \end{aligned}$$
 where $oldsymbol{\sigma}_k = egin{bmatrix} 0 \\ dots \\ 1 \\ 0 \end{bmatrix}$, $1 imes K$

if t is discrete, $any\ dimenstion = M$

$$\delta(t) = egin{bmatrix} 0 \ dots \ 1 \ 0 \end{bmatrix}, 1 imes M,$$

$$<\ln oldsymbol{\phi}, oldsymbol{x_n}> = \ln oldsymbol{\phi}^T \cdot oldsymbol{x_n}$$

$$=oldsymbol{\phi}^T\cdot\delta(t)$$

$$\begin{split} \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\phi}} &= \left\{ \sum_{n=1}^{N} [\boldsymbol{\gamma}_{n} \cdot \boldsymbol{x}_{n}^{T}] \right\}^{T} \odot \frac{1}{\boldsymbol{\phi}} - \sum_{t=1}^{M} u_{t} \cdot 1.\delta(t)^{T} \\ &= \left\{ \sum_{n=1}^{N} [\boldsymbol{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}] \right\} \odot \frac{1}{\boldsymbol{\phi}} - \begin{bmatrix} u_{1} & \cdots & u_{t} \cdots & u_{M} \\ u_{1} & \cdots & u_{t} \cdots & u_{M} \\ \vdots & & & \vdots \\ u_{1} & \cdots & u_{t} \cdots & u_{M} \end{bmatrix} = 0 \\ \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{A}} &= [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}] \odot \frac{1}{\boldsymbol{A}} - \sum_{k=1}^{K} v_{k} \cdot 1.\sigma_{k}^{T} \\ &= [\sum_{n=2}^{N} \boldsymbol{\xi}_{n}] \odot \frac{1}{\boldsymbol{A}} - \begin{bmatrix} v_{1} & \cdots & v_{k} \cdots & v_{K} \\ v_{1} & \cdots & v_{k} \cdots & v_{K} \\ \vdots & & & \vdots \\ v_{1} & \cdots & v_{k} \cdots & v_{K} \end{bmatrix} = 0 \\ \frac{\partial \boldsymbol{Q}}{\partial \boldsymbol{\pi}} &= \boldsymbol{\gamma}_{1} \odot \frac{1}{\boldsymbol{\pi}} - w_{1} \cdot 1. = 0 \end{split}$$

$$\frac{\left\{\sum_{n=1}^{N} [\mathbf{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}]\right\}_{ij}}{\boldsymbol{\phi}_{ij}} = u_{j}$$

$$\Rightarrow \sum_{i=1}^{K} \frac{\left\{\sum_{n=1}^{N} [\mathbf{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}]\right\}_{ij}}{u_{j}} = \sum_{i=1}^{K} \boldsymbol{\phi}_{ij} = 1$$

$$\Rightarrow u_{j} = \sum_{i=1}^{K} \left\{\sum_{n=1}^{N} [\mathbf{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}]\right\}_{ij}$$

$$\Rightarrow \boldsymbol{\phi}_{ij} = \frac{\left\{\sum_{n=1}^{N} [\mathbf{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}]\right\}_{ij}}{\sum_{i=1}^{K} \left\{\sum_{n=1}^{N} [\mathbf{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}]\right\}_{ij}}$$

$$\Rightarrow \boldsymbol{\phi} = \left\{\sum_{n=1}^{N} [\mathbf{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}]\right\} \odot \frac{1}{1 \cdot 1 \cdot 1^{T} \left\{\sum_{n=1}^{N} [\mathbf{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}]\right\}}$$

$$\Rightarrow \boldsymbol{\phi} = \left\{\sum_{n=1}^{N} [\mathbf{x}_{n} \cdot \boldsymbol{\gamma}_{n}^{T}]\right\} \odot \frac{1}{1 \cdot \left[\sum_{n=1}^{N} \boldsymbol{\gamma}_{n}\right]^{T}}$$

$$\frac{\left[\sum_{n=2}^{N} \boldsymbol{\xi}_{n}\right]_{ij}}{A_{ij}} = v_{j}$$

$$\Rightarrow \sum_{i=1}^{K} \frac{\left[\sum_{n=2}^{N} \boldsymbol{\xi}_{n}\right]_{ij}}{v_{j}} = \sum_{i=1}^{K} A_{ij} = 1$$

$$\Rightarrow v_{j} = \sum_{i=1}^{K} \sum_{n=2}^{N} \boldsymbol{\xi}_{n}\right]_{ij}$$

$$\Rightarrow \boldsymbol{A}_{ij} = \frac{\left[\sum_{n=2}^{N} \boldsymbol{\xi}_{n}\right]_{ij}}{\sum_{i=1}^{K} \left[\sum_{n=2}^{N} \boldsymbol{\xi}_{n}\right]_{ij}}$$

$$\Rightarrow \boldsymbol{A} = \left[\sum_{n=2}^{N} \boldsymbol{\xi}_{n}\right] \odot \frac{1}{1 \cdot 1 \cdot 1^{T} \left[\sum_{n=2}^{N} \boldsymbol{\xi}_{n}\right]}$$

$$\frac{\left[\boldsymbol{\gamma}_{1}\right]_{i}}{\boldsymbol{\pi}_{i}} = w_{1}$$

$$\Rightarrow \sum_{i=1}^{K} \frac{\left[\boldsymbol{\gamma}_{1}\right]_{i}}{w_{1}} = \sum_{i=1}^{K} \boldsymbol{\pi}_{i} = 1$$

$$\Rightarrow \boldsymbol{w}_{1} = \sum_{i=1}^{K} [\boldsymbol{\gamma}_{1}]_{i}$$

$$\Rightarrow \boldsymbol{\pi}_{1} = \frac{\left[\boldsymbol{\gamma}_{1}\right]_{i}}{\sum_{i=1}^{K} [\boldsymbol{\gamma}_{1}]_{i}}$$

$$\Rightarrow \boldsymbol{\pi}_{2} = \boldsymbol{\gamma}_{1} \odot \frac{1}{1 \cdot 1^{T} \boldsymbol{\gamma}_{1}}$$

To sum up

$$egin{aligned} oldsymbol{\phi} &= \left\{ \sum_{n=1}^N [oldsymbol{x_n} \cdot oldsymbol{\gamma_n}^T]
ight\} \odot rac{1.}{1.\left[\sum_{n=1}^N oldsymbol{\gamma_n}
ight]^T} \ oldsymbol{A} &= [\sum_{n=2}^N oldsymbol{\xi_n}] \odot rac{1.}{1.\cdot 1.^T [\sum_{n=2}^N oldsymbol{\xi_n}]} \ oldsymbol{\pi} &= oldsymbol{\gamma_1} \odot rac{1.}{1.\cdot 1.^T oldsymbol{\gamma_1}} \end{aligned}$$

How to calculate γ_n, ξ_n

$$m{\gamma}_n \equiv \left[\sum_{m{z}_n} p\left(m{z}_n | m{X}, m{ heta}^{
m old}
ight) m{z}_n
ight]$$
 为常数,是期望 $E[m{z}_n | m{X}, m{ heta}^{
m old}
ight]$,大小 $K imes 1$ $m{\xi}_n \equiv \left[\sum_{m{z}_{n-1}, m{z}_n} m{z}_n p\left(m{z}_{n-1}, m{z}_n | m{X}, m{ heta}^{
m old}
ight) m{z}_{n-1}^T
ight]$ 为常数,是期望 $E[m{z}_n m{z}_{n-1}^T | m{X}, m{ heta}^{
m old}
ight]$,大小

similarly get:

$$oldsymbol{\gamma_n} \equiv ec{p}(oldsymbol{z_n} | oldsymbol{X}, oldsymbol{ heta}^{
m old}) = egin{bmatrix} p(oldsymbol{z_n} | oldsymbol{X}, oldsymbol{ heta}^{
m old}) \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \ I \ | oldsymbol{z_n} = {
m state} \$$

so
$$\begin{split} & \gamma_n \equiv \vec{p}(z_n|X,\theta^{\text{old}}) \\ & = \vec{p}(z_n,X|\theta^{\text{old}}) \odot \frac{1.}{\vec{p}(X|\theta^{\text{old}})} \\ & = \vec{p}(z_n,x_1,\dots,x_n|\theta^{\text{old}}) \odot \vec{p}(x_{n+1},\dots,x_N|z_n,x_1,\dots,x_n,\theta^{\text{old}}) \odot \frac{1.}{\vec{p}(X|\theta^{\text{old}})} \\ & = \vec{p}(z_n,x_1,\dots,x_n|\theta^{\text{old}}) \odot \vec{p}(x_{n+1},\dots,x_N|z_n,\theta^{\text{old}}) \odot \frac{1.}{\vec{p}(X|\theta^{\text{old}})} \\ & = \vec{p}(z_n,x_1,\dots,x_n|\theta^{\text{old}}) \odot \vec{p}(x_{n+1},\dots,x_N|z_n,\theta^{\text{old}}) \odot \frac{1.}{\vec{p}(X|\theta^{\text{old}})} \\ & = \alpha_n \odot \beta_n \odot \frac{1.}{\vec{p}(X|\theta^{\text{old}})} \\ & = \vec{p}(z_{n-1},z_n|X,\theta^{\text{old}}) \odot \frac{1.}{\vec{p}(X|\theta^{\text{old}})} \\ & = \vec{p}(z_{n-1},x_1,\dots,x_{n-1}|\theta^{\text{old}}) \odot \vec{p}(z_{n-1},x_1,\dots,x_{n-1},\theta^{\text{old}}) \odot \vec{p}(z_n,x_n,\dots,x_N|z_{n-1},x_1,\dots,x_{n-1},\theta^{\text{old}}) \\ & \odot \vec{p}(z_n,x_n,\dots,x_N|z_{n-1},\theta^{\text{old}}) \odot \vec{p}(z_n,x_n,\dots,x_N|z_{n-1},\theta^{\text{old}}) \odot \vec{p}(z_n,x_n,\dots,x_N|z_{n-1},\theta^{\text{old}}) \odot \vec{p}(z_n|z_{n-1},\theta^{\text{old}}) \odot \vec{p}(x_n,\dots,x_N|z_{n-1},z_n,\theta^{\text{old}}) \odot \vec{p}(x_n,\dots,x_N|z_{n-1},z_n,\theta^{\text{old}}) \odot \vec{p}(x_n,\dots,x_N|z_{n-1},z_n,\theta^{\text{old}}) \odot \vec{p}(x_n,x_n,\dots,x_N|z_n,\theta^{\text{old}}) \odot \vec{p}(x_n,x_n,\dots,x_N|z_n,\theta^{\text{old}}) \odot \vec{p}(x_n,x_n,x_1,x_n,\theta^{\text{old}}) \odot \vec{p}(x_n,x_n,\theta^{\text{old}}) \odot \vec{p}(x_n,x_n,\theta^{\text{old}}) \odot \vec{p}(x_n,x_n,\theta^{\text{old}}) \odot \vec{p}(x_n|z_n,\theta^{\text{old}}) \odot \vec{p}(x_n|z_n,\theta^{\text{old$$

其中
$$p(\pmb{X}|\pmb{ heta}^{
m old}\,) = \sum_{Z_N} p(\pmb{z_N},\pmb{X}|\pmb{ heta}^{
m old}\,) = \sum \pmb{lpha_N} = 1.^T \cdot \pmb{lpha_N}$$
 is constant

$$egin{aligned} ec{p}(oldsymbol{X}|oldsymbol{ heta}^{
m old}) &= p(oldsymbol{X}|oldsymbol{ heta}^{
m old}) \cdot 1. \ p(oldsymbol{X}|oldsymbol{ heta}^{
m old}) &= 1.^T \cdot p(oldsymbol{z_N}, oldsymbol{X}|oldsymbol{ heta}^{
m old}) \ &= 1.^T \cdot oldsymbol{lpha_N} \end{aligned}$$

So if α_n, β_n is obtained, we could get γ_n, ξ_n

$$oldsymbol{lpha_n} \equiv ec{p}(oldsymbol{z_n}, oldsymbol{x_1}, \dots, oldsymbol{x_n} | oldsymbol{ heta}^{
m old} \,) \ oldsymbol{eta_n} \equiv ec{p}(oldsymbol{x_{n+1}}, \dots, oldsymbol{x_N} | oldsymbol{z_n}, oldsymbol{ heta}^{
m old} \,)$$

How to obtain α_n, β_n

for α_n , initial value:

$$egin{aligned} oldsymbol{lpha_1} & = ec{p}(oldsymbol{z_1}, oldsymbol{x_1} | oldsymbol{ heta}^{
m old} \) \ & = ec{p}(oldsymbol{x_1} | oldsymbol{z_1}, oldsymbol{ heta}^{
m old} \) \odot ec{p}(oldsymbol{z_1} | oldsymbol{ heta}^{
m old} \) \ & = [oldsymbol{\phi}^T \cdot oldsymbol{x_1}] \odot oldsymbol{\pi} \end{aligned}$$

Recurrence relation

$$\begin{split} &\boldsymbol{\alpha}_{\boldsymbol{n}} \equiv \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }}) \\ &= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\ &= \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \\ &= \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{\theta}^{\text{ old }}) \\ &= \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{ \vec{p}(\boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{\theta}^{\text{ old }}) \cdot 1. \right\} \\ &= \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{ \left[\vec{p}(\boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }}) \odot \vec{p}(\boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \right] \cdot 1. \right\} \\ &= \vec{p}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}}, \boldsymbol{\theta}^{\text{ old }}) \odot \left\{ \left[\vec{p}(\boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{\theta}^{\text{ old }})^{T} \right) \odot \vec{p}(\boldsymbol{z}_{\boldsymbol{n}} | \boldsymbol{z}_{\boldsymbol{n}-1}, \boldsymbol{\theta}^{\text{ old }}) \odot \left(1 \cdot \vec{p}(\boldsymbol{x}_{1}, \dots, \boldsymbol{x}_{\boldsymbol{n}-1} | \boldsymbol{z}^{\text{ old }})^{T} \right) \right] \cdot 1. \right\} \\ &= [\boldsymbol{\phi}^{T} \cdot \boldsymbol{x}_{\boldsymbol{n}}] \odot \left\{ \left[\boldsymbol{A} \odot \left(1 \cdot \boldsymbol{\alpha}_{\boldsymbol{n}-1}^{T} \right) \right] \cdot 1. \right\} \\ &= [\boldsymbol{\phi}^{T} \cdot \boldsymbol{x}_{\boldsymbol{n}}] \odot [\boldsymbol{A} \cdot \boldsymbol{\alpha}_{\boldsymbol{n}-1}] \end{split}$$

$$\boldsymbol{x}_{n}^{T} \cdot \boldsymbol{\phi} \cdot \boldsymbol{A} \cdot \boldsymbol{\alpha}_{n-1}$$

for β_n , last initial value:

$$oldsymbol{eta_N} = ec{p}(1|oldsymbol{z_N}, oldsymbol{ heta}^{
m old}) = 1$$

Recurrence relation

$$eta_{m{n}} \equiv ec{p}(m{x}_{n+1},\ldots,m{x}_N|m{z}_n,m{ heta}^{
m old}) = \left\{ec{p}(m{z}_{n+1},m{x}_{n+1},\ldots,m{x}_N|m{z}_n,m{ heta}^{
m old})^T \cdot 1.
ight\} = \left\{\left[ec{p}(m{z}_{n+1}|m{z}_n,m{ heta}^{
m old}) \odot ec{p}(m{x}_{n+1},\ldots,m{x}_N|m{z}_n,m{z}_{n+1},m{ heta}^{
m old})\right]^T \cdot 1.
ight\} = \left\{\left[ec{p}(m{z}_{n+1}|m{z}_n,m{ heta}^{
m old}) \odot \left(ec{p}(m{x}_{n+1},\ldots,m{x}_N|m{z}_{n+1},m{ heta}^{
m old}) \cdot 1.^T
ight)\right]^T \cdot 1.
ight\} = \left\{\left[ec{p}(m{z}_{n+1}|m{z}_n,m{ heta}^{
m old}) \odot \left(\left[ec{p}(m{x}_{n+1}|m{z}_{n+1},m{ heta}^{
m old}) \odot ec{p}(m{x}_{n+2},\ldots,m{x}_N|m{z}_{n+1},m{x}_{n+1},m{ heta}^{
m old})\right] \cdot 1.^T
ight)\right]^T \cdot 1.
ight\} = \left\{\left[m{A} \odot \left(\left[\left[m{\phi}^T \cdot m{x}_{n+1}\right] \odot m{ heta}_{n+1}\right] \cdot 1.^T
ight)\right]^T \cdot 1.
ight\} = m{A}^T \cdot \left(\left[m{\phi}^T \cdot m{x}_{n+1}\right] \odot m{ heta}_{n+1}
ight)$$

To sum up

$$egin{aligned} oldsymbol{lpha_1} &= [oldsymbol{\phi}^T \cdot oldsymbol{x_1}] \odot oldsymbol{\pi} \ oldsymbol{lpha_n} &= [oldsymbol{\phi}^T \cdot oldsymbol{x_n}] \odot [oldsymbol{A} \cdot oldsymbol{lpha_{n-1}}] \ oldsymbol{eta_N} &= oldsymbol{A}^T \cdot \left([oldsymbol{\phi}^T \cdot oldsymbol{x_{n+1}}] \odot oldsymbol{eta_{n+1}}
ight) \end{aligned}$$

Then

$$egin{aligned} oldsymbol{\gamma_n} &= oldsymbol{lpha_n} \odot oldsymbol{eta_n} \odot rac{1}{ec{p}(oldsymbol{X} | oldsymbol{ heta}
m{ old} \,)} \ &= rac{1}{1.^T \cdot oldsymbol{lpha_n}} igl[oldsymbol{lpha_n} \odot oldsymbol{eta_n} igr] \ oldsymbol{\xi_n} &= rac{1}{p(oldsymbol{X} | oldsymbol{ heta}
m{ old} \,)} iggl\{ igl[oldsymbol{\phi}^T \cdot oldsymbol{x_n} igr] \odot oldsymbol{eta_n} igr] \cdot oldsymbol{lpha_{n-1}} igr\} \odot oldsymbol{A} \ &= rac{1}{1.^T \cdot oldsymbol{lpha_n}} iggl\{ igl[oldsymbol{\phi}^T \cdot oldsymbol{x_n} igr] \odot oldsymbol{eta_n} igr] \cdot oldsymbol{lpha_{n-1}} igr\} \odot oldsymbol{A} \end{aligned}$$

because α_N, β_1 are so small

we divide $p(oldsymbol{X}|oldsymbol{ heta}^{
m old})$ into 2 part

$$p(\boldsymbol{X}|\boldsymbol{\theta}^{\text{ old }}) = p(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n|\boldsymbol{\theta}^{\text{ old }}) p(\boldsymbol{x}_{n+1}, \cdots, \boldsymbol{x}_N|\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n, \boldsymbol{\theta}^{\text{ old }})$$

define

$$egin{aligned} \hat{oldsymbol{lpha}}_{oldsymbol{n}} & = rac{1}{p(oldsymbol{x}_1, \cdots, oldsymbol{x}_n | oldsymbol{ heta} ext{ old })} oldsymbol{lpha}_{oldsymbol{n}} & = \left\{ \prod_{k=1}^n p(oldsymbol{x}_k | oldsymbol{x}_1, \cdots, oldsymbol{x}_{k-1}, oldsymbol{ heta} ext{ old })
ight\} oldsymbol{lpha}_{oldsymbol{n}} & = \left\{ \prod_{k=1}^n c_k
ight\} oldsymbol{lpha}_{oldsymbol{n}} & ext{ notice } & 1.^T \hat{oldsymbol{lpha}}_{oldsymbol{n}} & = \sum_{oldsymbol{z}_n} p(oldsymbol{z}_n | oldsymbol{x}_1, \cdots, oldsymbol{x}_n, oldsymbol{ heta} ext{ old }) = 1 \end{aligned}$$

the other way

$$egin{align*} c_{n} \hat{oldsymbol{lpha}}_{n} &= p(oldsymbol{x}_{n} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n-1}, oldsymbol{ heta}^{
m old}) egin{align*} oldsymbol{z}_{n} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n-1}, oldsymbol{ heta}^{
m old}) egin{align*} oldsymbol{z}_{n} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n-1}, oldsymbol{ heta}^{
m old}) \\ &= oldsymbol{p}(oldsymbol{z}_{n}, oldsymbol{z}_{n-1} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n-1}, oldsymbol{ heta}^{
m old}) \\ &= ig[oldsymbol{z}_{n}, oldsymbol{z}_{n-1} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n-1}, oldsymbol{ heta}^{
m old})^{T} ig] \odot oldsymbol{p}(oldsymbol{x}_{n} | oldsymbol{z}_{n} oldsymbol{ heta}^{
m old}) \\ &= ig[oldsymbol{p}(oldsymbol{z}_{n-1} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n-1}, oldsymbol{ heta}^{
m old}) ig] \odot oldsymbol{p}(oldsymbol{x}_{n} | oldsymbol{z}_{n} oldsymbol{ heta}^{
m old}) \\ &= ig[oldsymbol{\phi}(oldsymbol{z}_{n-1} | oldsymbol{x}_{1}, \cdots, oldsymbol{x}_{n-1}, oldsymbol{\theta}^{
m old}) ig] \odot oldsymbol{p}(oldsymbol{x}_{n} | oldsymbol{z}_{n} oldsymbol{\theta}^{
m old}) \\ &= oldsymbol{A} \cdot oldsymbol{\hat{\alpha}}_{n-1} ig] \odot oldsymbol{\phi}^{
m old} \cdot oldsymbol{x}_{n} \\ &= oldsymbol{\phi}^{
m old} \cdot oldsymbol{x}_{n} \odot oldsymbol{A} \cdot oldsymbol{\hat{\alpha}}_{n-1} ig] \end{split}$$

$$egin{aligned} \hat{m{lpha}}_{m{n}} &\equiv rac{1}{c_n} [m{\phi}^T \cdot m{x}_{m{n}}] \odot [m{A} \cdot \hat{m{lpha}}_{m{n-1}}] \ &= rac{1}{1.^T \cdot \left\{ [m{\phi}^T \cdot m{x}_{m{n}}] \odot [m{A} \cdot \hat{m{lpha}}_{m{n-1}}]
ight\}} [m{\phi}^T \cdot m{x}_{m{n}}] \odot [m{A} \cdot \hat{m{lpha}}_{m{n-1}}] \ \hat{m{lpha}}_{m{1}} &= rac{1}{1.^T \cdot \left\{ [m{\phi}^T \cdot m{x}_{m{n}}] \odot m{\pi}
ight\}} [m{\phi}^T \cdot m{x}_{m{n}}] \odot m{\pi} \end{aligned}$$

thesame

$$egin{aligned} egin{aligned} \hat{oldsymbol{eta}}_n & \equiv rac{oldsymbol{eta}_n}{oldsymbol{eta}_{n+1}} & = rac{rac{1}{p(x_{n+1},\cdots,x_N|x_1,\cdots,x_n,oldsymbol{ heta}^{
m old})}}{rac{1}{p(x_{n+2},\cdots,x_N|x_1,\cdots,x_{n+1},oldsymbol{ heta}^{
m old})}} \ & = rac{oldsymbol{eta}_n}{oldsymbol{eta}_{n+1}} & \cdot rac{p(oldsymbol{x}_1,\cdots,oldsymbol{x}_n|oldsymbol{ heta}^{
m old})}{p(oldsymbol{x}_1,\cdots,oldsymbol{x}_{n+1}|oldsymbol{ heta}^{
m old})} \ & = rac{oldsymbol{eta}_n}{oldsymbol{eta}_{n+1}} & \cdot rac{1}{p(oldsymbol{x}_{n+1}|oldsymbol{x}_1,\cdots,oldsymbol{x}_n,oldsymbol{ heta}^{
m old})} \ & = rac{oldsymbol{eta}_n}{oldsymbol{eta}_{n+1}} & \cdot rac{1}{c_{n+1}} \end{aligned}$$

the

$$egin{aligned} \hat{oldsymbol{eta}}_{oldsymbol{n}} &\equiv rac{1}{p(oldsymbol{x}_{n+1}, \cdots, oldsymbol{x}_N | oldsymbol{x}_1, \cdots, oldsymbol{x}_n, oldsymbol{ heta} ^{ ext{old}})} eta_{oldsymbol{n}} &= rac{1}{p(oldsymbol{x}_{n+1}, \cdots, oldsymbol{x}_N | oldsymbol{x}_1, \cdots, oldsymbol{x}_n, oldsymbol{ heta} ^{ ext{old}})} &= rac{1}{c_{n+1}} oldsymbol{A}^T \cdot \left([oldsymbol{\phi}^T \cdot oldsymbol{x}_{n+1}] \odot \hat{oldsymbol{eta}}_{n+1}
ight)} & \hat{oldsymbol{eta}}_{oldsymbol{N}} &= 1. \end{aligned}$$

then

$$egin{aligned} oldsymbol{\gamma_n} &= [\hat{oldsymbol{lpha}}_{oldsymbol{n}} \odot \hat{oldsymbol{eta}}_{oldsymbol{n}}] \ oldsymbol{\xi_n} &= rac{1}{c_n} \Bigg\{ \Big[[oldsymbol{\phi}^T \cdot oldsymbol{x}_n] \odot \hat{oldsymbol{eta}}_{oldsymbol{n}} \Big] \cdot \hat{oldsymbol{lpha}}_{oldsymbol{n}-1}^T \Bigg\} \odot oldsymbol{A} \end{aligned}$$

 α_N