$$\mathrm{KL}[q_1 q_2 q_3 \| p(123)]$$

$$= \langle \ln q_1 q_2 q_3 \rangle_{123} - \langle \ln p(123) \rangle_{123}$$

$$= \left\langle \ln \frac{q_1 q_2 q_3}{p(123)} \right\rangle_{123} = - \left\langle \ln \frac{p(123)}{q_1 q_2 q_3} \right\rangle_{123}$$

$$= -\left\langle \left\langle \ln \frac{p}{q_1 q_2 q_3} \right\rangle_{23} \right\rangle_1$$

$$= -\langle \langle \ln p \rangle_{23} - \langle \ln q_1 q_2 q_3 \rangle_{23} \rangle_1$$

$$= -\langle \langle \ln p \rangle_{23} - \{ \langle \ln q_1 \rangle_{23} + \langle \ln q_2 \rangle_{23} + \langle \ln q_3 \rangle_{23} \rangle_1$$

$$= -\langle \langle \ln p \rangle_{23} \underbrace{- \langle \ln q_1 \rangle_{23} \rangle_1}_{A} + \underbrace{\langle \langle \ln q_2 \rangle_{23} \rangle_1 + \langle \langle \ln q_3 \rangle_{23} \rangle_1}_{B}$$

A: $\int \underbrace{\ln q_1}_{\text{独立}} d(2.3)$

B: 2.3のみ ⇒ x_1をもたない const.

$$= -\langle \langle \ln p \rangle_{23} - \ln q_1 \rangle_1 + \mathrm{const}$$

$$= -\langle \ln \exp \langle \ln p \rangle_{23} - \ln q_1 \rangle_1 + \mathrm{Const}$$

$$= -\left\langle \ln \frac{\exp \langle \ln p \rangle_{23}}{q_1} \right\rangle_1 + \mathrm{Const}$$

$$= \mathrm{KL}[q_1 \| \exp \langle \ln p \rangle_{23}] + \mathrm{Const.}$$

$$KL[q_1 \| \exp\langle \ln p(123)\rangle_{23}]$$

$$= \int q_1 \ln \frac{\exp\langle \ln p(123)\rangle_{23}}{q_1} d\theta$$

ラグランジアン

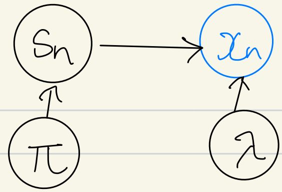
$$L = KL + \lambda\left(\int q(\theta)d\theta - 1\right)$$

$$\frac{\partial L}{\partial q(\theta)} = \frac{\partial}{\partial q(\theta)} \int q(\theta) \ln \exp\langle \sim \rangle - q(\theta) \ln q(\theta) d\theta + \lambda$$

$$= \underline{\{\ln \exp\langle \quad \rangle - \ln q(\theta)\}} + \lambda \quad = 0$$

$$\ln q(\theta) \propto \ln \exp\langle \ln p(123)\rangle_{23}$$

になる



① $p(\pi) = \mathrm{Dir}(\pi|\alpha)$

② $p(s_n|\pi) = \mathrm{Cat}(s_n|\pi)$

③ $p(\lambda_k) = \mathrm{Gam}(\lambda_k|a, b)$

④ $p(x_n|s_n, \lambda) = \prod_k p(x_n|\lambda_k)^{s_{n,k}}$ $\lambda = \{\lambda_1 \cdots \lambda_k\}$

同時分布 $p(X, S, \lambda, \pi) = p(X|S, \lambda)\, p(S|\pi)\, p(\lambda)\, p(\pi)$

$$= \left\{\prod_{n=1}^{N} \underbrace{p(x_n|s_n, \lambda)}_{\text{④}}\ \underbrace{p(s_n|\pi)}_{\text{②}}\right\}\left\{\prod_k \underbrace{p(\lambda_k)}_{\text{③}}\right\}\underbrace{p(\pi)}_{\text{①}}$$

④ $s_{n,k}$ 時点n、クラスkの潜在変数. $s_{n,k} \in \{0, 1\}$

$p(x_n|s_n, \lambda) = \prod_k \mathrm{Poisson}(x_n|\lambda_k)^{s_{n,k}}$

注、$\sum_k s_{n,k} = 1$ であり、クラスがひとつだけ実現

変分推論

$p(S\lambda\pi|X) \propto p(XS\lambda\pi)$ を
事後分布

$q(S)q(\lambda,\pi)$ で近似

Sに注目する

$$\ln q(S) = \langle \ln p(X|S,\lambda)p(S|\pi)p(\lambda)p(\pi)\rangle_{q(\lambda,\pi)} + \mathrm{Const}$$

$$= \underbrace{\langle \ln p(X|S,\lambda)\rangle_{q(\lambda,\pi)} + \langle \ln p(S|\pi)\rangle_{q(\lambda,\pi)}}_{A} + \underbrace{\langle \ln p(\lambda)\rangle_{q(\lambda,\pi)} + \langle \ln p(\pi)\rangle_{q(\lambda,\pi)}}_{B}$$

B は $\langle \ln p(\pi)\rangle_{q(\lambda,\pi)} + \langle \ln p(\lambda)\rangle_{q(\lambda,\pi)}$ で、Sに無関係な定数
→ Const にする

A は $\langle \ln p(X|S,\lambda)\rangle_{q(\lambda,\pi)}$ は、πが存在しない→無条件で周辺化
(πを消去)

→ $\langle \ln p(X|S,\lambda)\rangle_{q(\lambda)}$

同様に、$\langle \ln p(S|\pi)\rangle_{q(\lambda,\pi)}$ は λが存在しない。

→ $\langle \ln p(S|\pi)\rangle_{q(\pi)}$

以上を整理すると

$$\ln q(S) = \langle \ln p(X|S,\lambda)\rangle_{q(\lambda)} + \langle \ln p(S|\pi)\rangle_{q(\pi)} + \sim\sim$$

→

対数尤度の計算（$\ln q(S)$）

$$\ln p(x_n|s_n,\lambda) = \sum_k s_{n,k} \ln \mathrm{Poisson}(x_n|\lambda_k)$$
$$= \sum_k s_{n,k}(x_n \ln\lambda_k - \underline{\ln x_n!} - \lambda_k)$$

$$\ln p(s_n|\pi) = \ln \mathrm{Cat}(s_n|\pi) \left(= \prod \pi_k^{s_{n,k}}\right)$$
$$= \sum_k s_{n,k} \ln \pi_k$$

これらが iid
つまり $\sum_{n=1}^{N}$ ◯
と書けばOK

を利用すると

A
$$\langle \ln p(X|S,\lambda)\rangle_{q(\lambda)}$$

λ含まないので const へ.

$$= \left\langle \sum_n \sum_k s_{n,k}(x_n \ln\lambda_k - \boxed{\ln x_n!} - \lambda_k)\right\rangle_{q(\lambda)}$$

$$= \sum_n \sum_k s_{n,k}(x_n\langle \ln\lambda_k\rangle - \langle\lambda_k\rangle) + \mathrm{const.}$$

B
$$\langle \ln p(S|\pi)\rangle_{q(\pi)}$$
$$= \left\langle \sum_n \sum_k s_{n,k} \ln\pi_k \right\rangle_{q\pi}$$
$$= \sum_n \sum_k s_{n,k}\langle \ln\pi_k\rangle$$

A+B
全部足す

$$\sum_n \sum_k s_{n,k}\underline{(x_n\langle\ln\lambda_k\rangle - \langle\lambda_k\rangle + \langle\ln\pi_k\rangle)}$$

$\ln \exp(x_n$ 〜〜〜 $)$ と見ると $\sum_k s_{n,k} \ln(\sim)$
→つまりカテゴリ分布の尤度の形になっている.

対数尤度 $\ln(q(\lambda,\pi))$

$$\ln q(\lambda,\pi) = \langle \ln p(X|S,\lambda)\, p(S|\pi)\, p(\lambda)\, p(\pi) \rangle_{q(S)}$$

$$= \underbrace{\langle \ln p(X|S,\lambda)\rangle_{(S)} + \langle \ln p(S|\pi)\rangle_{(S)}}_{A} + \underbrace{\langle \ln p(\lambda)\rangle_{(S)} + \langle \ln p(\pi)\rangle_{(S)}}_{B}$$

Bは $\langle \ln p(\lambda)\rangle_{(S)}$ と、変数に無関なもので積分→外に出る

$$\left(\int f(y)\, g(x)\, dx = f(y) \int g(x)\, dx \right)$$

$$\langle \ln p(\lambda)\rangle_{(S)} = \ln p(\lambda)$$

$$\langle \ln p(\pi)\rangle_{(S)} = \ln p(\pi)$$

AとB整理すると

$$\underbrace{\langle \ln p(X|S,\lambda)\rangle_{(S)} + \ln p(\lambda)}_{\lambda だけ} + \underbrace{\langle \ln p(S|\pi)\rangle_{S} + \ln p(\pi)}_{\pi だけ}$$

$$\ln q(\lambda,\pi) = f(\lambda) + g(\pi) \Rightarrow q(\lambda,\pi) \propto \underbrace{e^{f(\lambda)}\, e^{g(\pi)}}$$

→独立

つまり別々に考えればOK.

$$\ln q(\lambda) = \langle \ln p(X|S,\lambda)\rangle_{(S)} + \ln p(\pi)$$

$$= \left\langle \sum_n \sum_k s_{n,k}(x_n \ln \lambda_k - \ln x_n! - \lambda_k) \right\rangle + \sum_k \{(a-1)\ln \lambda_k - b\lambda_k + C\}$$

$$= \sum_k \left\{ \underline{\left(\sum_n \langle s_{n,k}\rangle x_n + a - 1\right) \ln \lambda_k - \left(\sum_n \langle s_{n,k}\rangle + b\right) \lambda_k} \right\}$$

$$\sum_k (\bigcirc - 1)\ln \lambda_k - \sum_k \square\, \lambda_k$$ の形

クラスごとにガンマ分布

$$\ln q(\pi) = \langle \ln p(S|\pi) \rangle + \ln p(\pi)$$

$$= \left\langle \sum_n \sum_k s_{n,k} \ln \pi_k \right\rangle + \sum_k (\alpha_k - 1) \ln \pi_k + c.$$

$$= \sum_n \sum_k \langle s_{n,k} \rangle \ln \pi_k + \sum_k (\alpha_k - 1) \ln \pi_k$$

$$= \sum_k \underline{\left(\sum_n \langle s_{n,k} \rangle + \alpha_k - 1 \right) \ln \pi_k} + c$$

$\sum \square \ln \pi$ の形 → クラスごとにディリクレ分布

以上、書籍 P60〜61. (2.59) や (2.60) を使って.

$$\langle \lambda_k \rangle = \frac{\sum \langle s_{n,k} \rangle x_n + a}{\sum_n \langle s_{n,k} \rangle + b}$$

$$\langle \ln \lambda_k \rangle = \psi\left(\sum_n \langle s_{n,k} \rangle x_n + a \right) - \ln\left(\sum_n \langle s_{n,k} \rangle + b \right)$$

$$\langle \ln \pi_k \rangle = \psi\left(\sum_n \langle s_{n,k} \rangle + \alpha_k \right) - \psi\left(\sum_k \left(\sum_n \langle s_{n,k} \rangle + \alpha_k \right) \right)$$

※ $\psi(\cdot)$: ディガンマ

$$\langle s_{n,k} \rangle \propto \exp\{ x_n \langle \ln \lambda_k \rangle - \langle \lambda_k \rangle + \langle \ln \pi_k \rangle \} = \eta_{n,k}$$

これを n 時点ごとに. k で 正規化 すれば $\langle s_{n,k} \rangle$ になる

$$= \frac{\eta_k}{\sum_k \eta_{n,k}}$$

HMM

同時分布

$$p(X, S, \lambda, \pi, A)$$

$$= p(X|S, \lambda)\, p(S|\pi, A)\, p(\lambda)\, p(\pi)\, p(A) \quad \text{---} (※)$$

$$= (\lambda)(\pi)(A)(x_1|S_1, \lambda)(S_1|\pi) \prod_{n=2}^{N} (x_n|S_n, \lambda)(S_n|S_{n-1}, A)$$

変分推論

$$\ln(\lambda, \pi, A) = \langle \ln p(X, S, \lambda, \pi, A) \rangle_{(S)}$$

$$= \langle \ln(\lambda) \rangle_{(S)} + \langle \ln(\pi) \rangle_{(S)} + \langle \ln(A) \rangle_{(S)} + \sum_n \langle \ln(x_n|S_n, \lambda) \rangle_S + \langle \ln(S|\pi, A) \rangle_S$$

λだけ、πだけ、Aだけ で分解可能 → 独立

$$\ln q(\lambda) = \ln(\lambda) + \sum_{n=2} \langle \ln(x_n|S_n, \lambda) \rangle_S$$

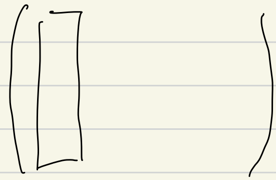
$$\ln q(\pi) = \ln(\pi) + \langle \ln(S_1|\pi) \rangle_S$$

$$\ln q(A) = \ln(A) + \sum_{n=2} \langle \ln(S_n|S_{n-1}, A) \rangle_S$$

……だが、実は完全分解と同じ。以降不要。

forward

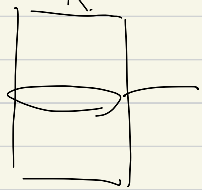
S_{n-1}

$A = S_n$ 

列和 $=1$

$\sum_{ij} a_{ij} = K$

K

$f =$ n  $f(S_n) = (1 \times K)$

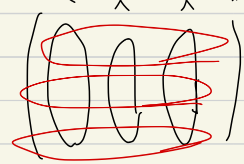
$p(S_n \mid S_{n-1}) \times f(S_{n-1})$

$\underline{A \quad \mathrm{Diag}(f(S_{n-1}))}$

S_{n-1}

$f(S_{n-1})$ $f(S_{n-1})$ $f(S_{n-1})$

$\times$ $\times$ $\times$

S_n 

縦和$=1$

これを行和計算したら $p(S_n)$

$p(x_n \mid S_n) \times p(S_n)$

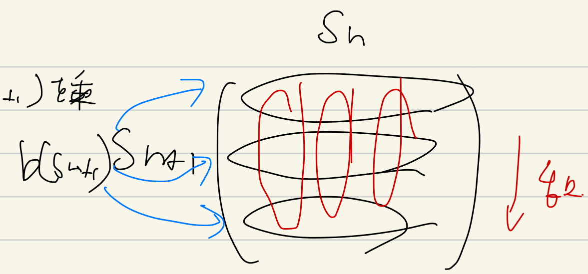
行ベクトルとする

$$\underbrace{f(S_n)}_{\text{ベクトル}} = \underbrace{\hat{p}(x_n \mid S_n)}_{\text{ベクトル}} \sum_{S_{n-1}} \underbrace{\hat{p}(S_n \mid S_{n-1})}_{\text{行列}(A)} \underbrace{f(S_{n-1})}_{\text{ベクトル}}$$

$$\tilde{p}(x_n \mid S_n) \odot \left\{ \underbrace{\left(A \ \mathrm{Diag}\, f(S_{n-1}) \right)}_{A\text{の各列を}f(S_{n-1})\text{倍する}} \underbrace{1_{(K \times 1)}}_{\text{行和計算}} \right\}$$

backward.

S_{n+1} の各行に $b(S_{n+1})$ を乗



$\mathrm{Diag}\{b(S_{n+1})\}\ A$

次に各行に $\tilde{p}(x_{n+1}|S_{n+1})$ をかける.

最後に S_{n+1} を列和で計算

$$\sum_{n+1} \tilde{p}(x_{n+1}|S_{n+1})\, \tilde{p}(S_{n+1}|S_n)\, b(S_{n+1})$$

$$b(n) = \mathbb{1}_{1\times 4}\ \mathrm{Diag}\{\tilde{p}(x_{n+1}|S_{n+1}) \odot b(S_{n+1})\}\ A.$$

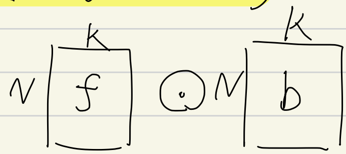
$$(1, 1, 1)\begin{pmatrix} b_1 P_1 & & \\ & b_2 P_2 & \\ & & b_3 P_3 \end{pmatrix} A.$$

クラスごとの $b(S_{n+1})\ P(x_{n+1}|S_{n+1})$

Forwardと計算の方向が違うので注意する

$q(s_n)$ と $q(s_{n-1}, s_n)$

$q(s_n) \propto f(s_n)\, b(s_n)$

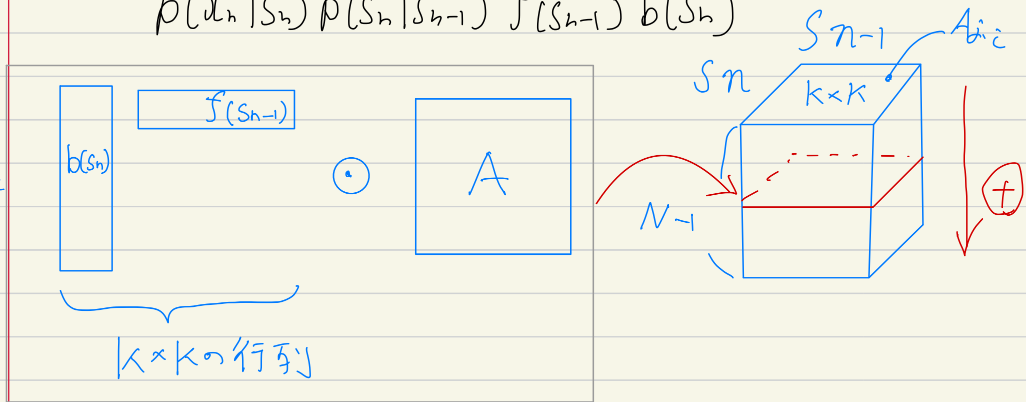


で OK

$q(s_{n-1}, s_n)$

$\tilde{p}(x_n|s_n)\,\hat{p}(s_n|s_{n-1})\, f(s_{n-1})\, b(s_n)$

bfa := 行列



これの各行に $p(x_n|s_n)$ をかける。

bfa $\mathrm{Diag}(p(x_n|s_n))$

$K \times K \times N-1$ の配列を $(N-1)$ 方向に和