Ex. 1

1. $$p_\theta(X) = \int p_\theta(X,S)\, dS = \int \prod_{t=1}^{T} p_\theta(x_t|s_t)\, p_\theta(s_t)\, dS =$$

$p(X;\theta_{k-1})$

$$= \prod_{t=1}^{T} \int p_\theta(x_t|s_t)\, p_\theta(s_t)\, ds_t = \prod_{t=1}^{T} p_\theta(x_t) =: \prod_{t=1}^{T} p(x_t;\theta)$$

2. $$J(\theta) = \int \log\big(p(X,S;\theta)\big)\, p(S\,|\,X;\theta_{k-1})\, dS$$

$$= \int \log\Big(\prod_{t=1}^{T} \underbrace{p(x_t|s_t;\theta)\cdot p(s_t;\theta)}_{p(x_t,s_t;\theta)}\Big)\cdot \frac{p(X,S;\theta_{k-1})}{p(X;\theta_{k-1})}\, ds_1\dots ds_T$$

$$= \int \sum_{t=1}^{T} \big(\log p(x_t,s_t;\theta)\big)\cdot \frac{\prod_{\tau=1}^{T} p(x_\tau|s_\tau;\theta_{k-1})\cdot p(s_\tau;\theta_{k-1})}{\prod_{\tau=1}^{T} p(x_\tau;\theta_{k-1})}\, ds_1\dots ds_T$$

$$= \int \sum_{t=1}^{T} \big(\log p(x_t,s_t;\theta)\big)\cdot \prod_{\tau=1}^{T} \underbrace{\frac{p(x_\tau|s_\tau;\theta_{k-1})\cdot p(s_\tau;\theta_{k-1})}{p(x_\tau;\theta_{k-1})}}_{p(s_\tau|x_\tau;\theta_{k-1})}\, ds_1\dots ds_T$$

$$= \sum_{t=1}^{T} \Big\{\int \log p(x_t,s_t;\theta)\cdot \prod_{\tau=1}^{T} p(s_\tau|x_\tau;\theta_{k-1})\, ds_1\dots ds_T\Big\}$$

$$= \sum_{t=1}^{T} \Big\{ \underbrace{\int \prod_{\substack{\tau=1\\ \tau\neq t}}^{T} p(s_\tau|x_\tau;\theta_{k-1})\, ds_\tau}_{\prod_{\tau\neq t} \underbrace{\int p(s_\tau|x_\tau;\theta_{k-1})\, ds_\tau}_{=1}} \cdot \int \log p(x_t,s_t;\theta)\cdot p(s_t|x_t;\theta_{k-1})\, ds_t$$

$$= \sum_{t=1}^{T} \int \log p(x_t,s_t;\theta)\cdot p(s_t|x_t;\theta_{k-1})\, ds_t$$

For discrete variables the calculation is analogue, when replacing the integral over S with a sum over S.
This expression has a nice interpretation: $\log p(x_t,s_t;\theta)$, which would give the complete log-likelihood when summed-up, is replaced by an estimate, namely the conditional expectation.

3. $p(r(t) \mid x(t); \underset{\uparrow}{\Theta}) = \frac{p(r(t), x(t); \Theta)}{p(x(t); \Theta)} =$

short notation for
all parameters
$\mu_c, C_c, c = 1 \dots, C$

$$= \frac{p(r(t), x(t); \Theta)}{\sum_{r(t)=1}^{C} p(r(t), x(t); \Theta)} = \frac{q_{tc}}{\sum_{c=1}^{C} q_{tc}} = q_{tc}^{*}$$

Ex. 2

For 1. and 2. $x_i := u_i$, $x := u$.

1. $$E(x_i) = \sum_{x_i=0,1} x_i\, p(x_i) = 0\cdot\mu_i^0(1-\mu_i)^1 + 1\cdot\mu_i^1(1-\mu_i)^0 = \mu_i$$

$$V(x_i) = E(x_i^2) - E(x_i)^2 = \sum_{x_i=0,1} x_i^2\, p(x_i) - \mu_i^2 = 0^2\cdot\mu_i^0(1-\mu_i)^1 + 1^2\mu_i^1(1-\mu_i)^0 - \mu_i^2 =$$

$$= \mu_i - \mu_i^2 = \mu_i(1-\mu_i)$$

2. $$E(x) = E\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} E(x_1) \\ \vdots \\ E(x_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \mu$$

$$\operatorname{cov}(x_i, x_j) = E(x_i x_j) - E(x_i)E(x_j) \underset{\substack{x_i, x_j \\ \text{indep.}}}{=} E(x_i)E(x_j) - E(x_i)E(x_j) = 0$$

$$\operatorname{cov}(x) = \begin{pmatrix} V(x_1) & \operatorname{cov}(x_1,x_2) & & \\ \operatorname{cov}(x_1,x_2) & V(x_2) & & \\ & & \ddots & \\ & & & V(x_n) \end{pmatrix} = \operatorname{diag}(\mu_i(1-\mu_i)) =: D_c$$

3. $$E(x) = \sum_{x\in\{0,1\}^n} x\, q(x) = \sum_{x\in\{0,1\}^n} x \sum_{c=1}^{C} \pi_c\, p(x;\mu_c) = \sum_{c=1}^{C} \pi_c \underbrace{\sum_{x\in\{0,1\}^n} x\, p(x;\mu_c)}_{=\,\mu_c \text{ (from 2.)}}$$

$$= \sum_{c=1}^{C} \pi_c \mu_c$$

First note that $$q(x_i) = \sum_{\substack{x_k \\ k=1,\dots,n \\ k\neq i}} q(x_1,\dots,x_n) = \sum_{\substack{x_k \\ k=1,\dots,n \\ k\neq i}} \sum_{c=1}^{C} \pi_c\, p(x;\mu_c) =$$

$$= \sum_{c=1}^{C} \pi_c \underbrace{\sum_{\substack{x_k \\ k=1,\dots,n \\ k\neq i}} p(x;\mu_c)}_{p(x_i;\mu_c)} = \sum_{c=1}^{C} \pi_c \cdot p(x_i;\mu_c)$$

$$E(x_i^2) = \sum_{x_i=0}^{1} x_i^2\, q(x_i) = \sum_{x_i=0}^{1} x_i^2 \sum_{c=1}^{C} \pi_c\, p(x_i;\mu_c) = \sum_{c=1}^{C} \pi_c \underbrace{\sum_{x_i=0}^{1} x_i^2\, \mu_{ci}^{x_i}(1-\mu_{ci})^{(1-x_i)}}_{\mu_{ci}}$$

$$= \sum \pi_c \mu_{ci}$$

$$E(x_i x_j) = \sum_{x_i=0}^{1} \sum_{x_j=0}^{1} x_i x_j \sum_{c=1}^{C} \pi_c \underbrace{p(x_i, x_j; \mu_c)}_{p(x_i;\mu_c)\cdot p(x_j;\mu_c)}$$

$$= \sum_{c=1}^{C} \pi_c \sum_{x_i=0}^{1} \sum_{x_j=0}^{1} x_i\, p(x_i;\mu_c) \cdot x_j\, p(x_j;\mu_c)$$

$$= \sum_{c=1}^{C} \pi_c \underbrace{\Big(\sum_{x_i=0}^{1} x_i\, p(x_i;\mu_c)\Big)}_{\mu_{ci}} \underbrace{\Big(\sum_{x_j=0}^{1} x_j\, p(x_j;\mu_c)\Big)}_{\mu_{cj}}$$

$$= \sum_{c=1}^{C} \pi_c\, \mu_{ci}\, \mu_{cj}$$

$$\mathrm{Cov}(x) = E(xx^T) - E(x)E(x)^T$$

$$= \sum_{c=1}^{C} \pi_c \begin{pmatrix} \mu_{c1} & \mu_{c1}\mu_{c2} & \cdots & \\ \mu_{c1}\mu_{c2} & \mu_{c2} & & \\ \vdots & & \ddots & \\ & & & \mu_{cn} \end{pmatrix} - \left\{ \sum \pi_c \begin{pmatrix} \mu_{c1} \\ \vdots \\ \mu_{cn} \end{pmatrix} \right\} \cdot \left\{ \sum \pi_c \,(\mu_{c1} \cdots \mu_{cn}) \right\}$$

From this we can see that the covariance matrix is not diagonal, that shows that the mixture distribution is a richer distribution than a single multivariate Bernoulli distribution.

4. $$\log\Big(\prod_{t=1}^{T} q(x(t); \mu_c, \pi_c, c=1,\dots,C)\Big) = \sum_{t=1}^{T} \log q(x(t); \mu_c, \pi_c) =$$

$$= \sum_{t=1}^{T} \log \sum_{c=1}^{C} \pi_c\, p(x(t); \mu_c)$$

$$= \sum_{t=1}^{T} \log \Big\{ \sum_{c=1}^{C} \pi_c \prod_{i=1}^{n} p(x_i(t); \mu_{ci}) \Big\}$$

5. $$\log\Big(\prod_{t=1}^{T} q(x(t), r(t); \mu_{r(t)})\Big) = \sum_{t=1}^{T} \log\big(q(x(t), r(t); \mu_{r(t)})\big)$$

$$= \sum_{t=1}^{T} \log\big\{ \pi_{r(t)}\, p(x(t); \mu_{r(t)}) \big\} = \sum_{t=1}^{T} \Big\{ \log \pi_{r(t)} + \log \prod_{i=1}^{n} \underbrace{p(x_i(t); \mu_{r(t),i})}_{\mu_{r(t),i}^{x_i(t)} \cdot (1-\mu_{r(t),i})^{(1-x_i(t))}} \Big\}$$

$$= \sum_{t=1}^{T} \log \pi_{r(t)} + \sum_{t=1}^{T} \sum_{i=1}^{n} \big\{ x_i(t) \log \mu_{r(t),i} + (1-x_i(t)) \log(1-\mu_{r(t),i}) \big\} =: \ell$$

6. $$q(r(t)=c \mid x(t)) = \frac{q(r(t), x(t))}{q(x(t))} = \frac{\pi_{r(t)}\, p(x(t);\mu_{r(t)})}{\sum_{c=1}^{C} q(x(t), r(t)=c)}$$

$$= \frac{\overset{\pi_c}{\pi_{r(t)}}\; p(x(t);\overset{\mu_c}{\mu_{r(t)}})}{\sum_{k=1}^{C} \pi_k\, p(x(t);\mu_k)}$$

(This is equivalent to Eq. 10.16 in the lecture handout for the Gaussian mixture!)

7. $$E(\ell(\mu_c,\pi_c)) \overset{\text{math ex. 1}}{=} \sum_{t=1}^{T}\sum_{c=1}^{C} \underbrace{q(r(t)=c \mid x(t))}_{q^*_{t,c}} \cdot \log\left(q(x(t), r(t);\mu_{r(t)})\right)$$

$$= \sum_{c=1}^{C}\sum_{t=1}^{T} q^*_{tc} \cdot \log \pi_c + \sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{i=1}^{n} q^*_{t,c} \cdot \left(x_i(t)\log\mu_{ci} + (1-x_i(t))\log(1-\mu_{ci})\right)$$

$$=: J$$

8. $$\frac{\partial J}{\partial \mu_{ci}} = 0 + \sum_{t=1}^{T} q^*_{tc}\left(x_i(t)\cdot\frac{1}{\mu_{ci}} - (1-x_i(t))\frac{1}{1-\mu_{ci}}\right)$$

$$= \frac{1}{\mu_{ci}}\sum_{t=1}^{T} q^*_{tc}\, x_i(t) - \frac{1}{1-\mu_{ci}}\sum_{t=1}^{T} q^*_{tc}\,(1-x_i(t)) \overset{!}{=} 0$$

$$\Rightarrow (1-\mu_{ci})\sum_{t=1}^{T} q^*_{tc}\, x_i(t) = \mu_{ci}\left\{\sum_{t=1}^{T} q^*_{tc} - \sum_{t=1}^{T} q^*_{tc}\, x_i(t)\right\}$$

$$\sum_{t=1}^{T} q^*_{tc}\, x_i(t) - \cancel{\mu_{ci}\sum_{t=1}^{T} q^*_{tc}\, x_i(t)} = \mu_{ci}\sum_{t=1}^{T} q^*_{tc} - \cancel{\mu_{ci}\sum_{t=1}^{T} q^*_{tc}\, x_i(t)}$$

$$\Rightarrow \mu_{ci} = \frac{\sum_{t=1}^{T} q^*_{tc}\, x_i(t)}{\sum_{t=1}^{T} q^*_{tc}}$$

use $N_c := \sum_{t=1}^{T} q^*_{tc}$ and we get $\mu_{ci} = \frac{1}{N_c}\sum_{t=1}^{T} q^*_{tc}\, x_i(t)$

$$\Rightarrow \boxed{\mu_c = \frac{1}{N_c}\sum_{t=1}^{T} q^*_{tc}\, x(t)}$$

(This is the same update rule as for Gaussian mixtures, compare Eq. 10.17)

9. $\tilde{J} = J + \lambda \left(1 - \sum_{k=1}^{c} \pi_k \right)$

$$\frac{\partial \tilde{J}}{\partial \pi_k} = \underbrace{\sum_{t=1}^{T} q_{tk}^*}_{N_k} \frac{1}{\pi_k} - \lambda = \frac{N_k}{\pi_k} - \lambda \overset{!}{=} 0$$

$$\Rightarrow \quad \pi_k = \frac{N_k}{\lambda}$$

use the constraint to calculate λ:

$$1 = \sum_{k=1}^{c} \pi_k = \sum_{k=1}^{c} \frac{N_k}{\lambda} = \frac{1}{\lambda} \sum_{k=1}^{c} N_k$$

$$\Rightarrow \quad \lambda = \sum_{k=1}^{c} N_k = \sum_{k=1}^{c} \sum_{t=1}^{T} q_{tk}^* = \sum_{t=1}^{T} \sum_{k=1}^{c} \frac{\pi_k\, p(x(t); \mu_k)}{\sum_{\ell=1}^{c} \pi_\ell\, p(x(t); \mu_\ell)}$$

$$= \sum_{t=1}^{T} \frac{\sum_{k=1}^{c} \pi_k\, p(x(t); \mu_k)}{\sum_{\ell=1}^{c} \pi_\ell\, p(x(t); \mu_\ell)} = \sum_{t=1}^{T} 1 = T$$

$$\Rightarrow \quad \boxed{\pi_k = \frac{N_k}{T}}$$

(This is the same as the update rule for the Gaussian mixture in Eq. 10.19)