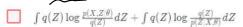
1. Recall the derivation of EM algorithm. In our notation, X is observable data, Z is latent variable and  $\theta$  is a vector of model parameters. We introduced q(Z) — an arbitrary distribution over the latent variable. Choose the correct expressions for the marginal log-likelihood  $\log p(X\mid\theta)$ :



7

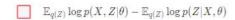
This should be selected



## Correct

Z is integrated out:

 $\log \int p(X, Z|\theta) dZ = \log p(X|\theta)$ 



7

This should be selected

 $\int q(Z) \log p(X|\theta) dZ$ 



This should be selected

2. In EM algorithm, we maximize variational lower bound  $\mathcal{L}(q,\theta) = \log p(X|\theta) - \mathrm{KL}(q||p) \text{ with respect to } q \text{ (E-step) and } \theta \text{ (M-step) iteratively.}$  Why is the maximization of lower bound on E-step equivalent to minimization of KL divergence?











## This should not be selected

Revise E-step details video



6. Imagine that you want to pat your friend's cat Becky. Cats are really random creatures.



Becky might get grumpy and scratch you with probability p or curl up and start purring (with prob. 1-p). You don't know Becky well yet, so you estimate prior on p to be distributed as Beta(2,2). Within one evening, Becky has scratched you  $\underline{6}$  times and only 2 times she purred. What will be the parameters for posterior distribution over p? What is the MAP-estimate for p?

Enter your answers separated by comma: e.g. if you think that correct answer is Beta(1,0.2) and MAP is 3, you should enter 1,0.2,3. Express real numbers as decimals with dot as delimiter.

8,4,0.7

Correct Response

1. Recall the derivation of EM algorithm. In our notation, X is observable data, Z is latent variable and  $\theta$  is a vector of model parameters. We introduced q(Z) — an arbitrary distribution over the latent variable. Choose the correct expressions for the marginal log-likelihood  $\log p(X\mid\theta)$ :

Correct

 $\int q(Z) \log \frac{p(X,Z|\theta)}{q(Z)} dZ + \int q(Z) \log \frac{q(Z)}{p(Z|X,\theta)} dZ =$ 

 $\int q(Z) \log p(X, Z|\theta) dZ - \int q(Z) \log q(Z) dZ +$ 

 $+ \int q(Z) \log q(Z) dZ - \int q(Z) \log p(Z|X, \theta) dZ =$ 

 $=\int q(Z)\log rac{p(X,Z|\theta)}{p(Z|X,\theta)}dZ = \int q(Z)\log p(X|\theta)dZ = \log p(X|\theta)$ 

" p(x 210)/ p(x,13)

 $\log \int p(X, Z|\theta)dZ$ 

Correct

Z is integrated out:

 $\log \int p(X, Z|\theta) dZ = \log p(X|\theta)$ 

 $\mathbb{E}_{q(Z)} \log p(X, Z|\theta) - \mathbb{E}_{q(Z)} \log p(Z|X, \theta)$ 

Correct

 $\mathbb{E}_{q(Z)} \log p(X, Z|\theta) - \mathbb{E}_{q(Z)} \log p(Z|X, \theta) =$ 

 $= \mathbb{E}_{q(Z)} \log \tfrac{p(X,Z|\theta)}{p(Z|X,\theta)} = \mathbb{E}_{q(Z)} \log p(X|\theta) = \log p(X|\theta)$ 

=1(2)(0)/ p(x (0)

 $\boxed{\quad \ \ } \int q(Z)\log p(X|\theta)dZ$ 

Correct

 $\log p(X|\theta)$  does not depend on Z.

 $\int q(Z) \log p(X|\theta) dZ = \log p(X|\theta)$ 

$$\int_{1}^{2} (y(\theta)) = \int_{1}^{2} \left( \frac{1}{1-\theta} \right)^{\frac{1}{1-\theta}} = \frac{1}{2} \frac$$

Calculable

(Node , not mean

Mode of Beta dist

$$\frac{\sqrt{-1}}{\sqrt{+\beta-2}} = \frac{8-1}{8-4-2} = \frac{7}{16}$$