- f. Derivation of Conditional VAE c Reference; EM algorithm L13 P. 23)
- · The chain rule of probability

log p(X; 0) = log p(X, Z; 0) - log p(Z|X; 0)

Given c log p(Xlc; 0) = log p(X.Zlc; 0) - log p(Z|X,c; 0)

· We next introduce an arbitrary distribution q(Zlc) on both sides and integrate over Z.

Sq(Zlc) log p(Xlc; 0) dZ = Sq(Zlc) log p(X,Zlc; 0) dZ - Sq(Zlc) log p(ZIX,c;0) dZ

= Sq(Zlc) logp(X,Zlc; 0) dZ - Sq(Zlc) logq(Zlc) dZ

+ Sq(Zle) log q(Zle) dZ - Sq(Zlc) log p(ZIX,c;0) dZ

= L(X,c,q,0) + KL(q(Z|d) || p(Z|X,c;0))

where

kl (q(Z1c) 11 p(Z1X,c;0)) = \(q(Z1c) log p(X,Z1c;0) dZ - \sq(Z1c) log q(Z1c) dZ \)

* Since the KL divergence is non-negative, KL (QIIP) >0, it follows that

log p(X1c; 0) > L (X, c, q, 0)

with equality if and only if

q(Z(c) = p(Z(X,c; 0)

In other words, L(X, c, q, 0) is a lower bound on log p(X/c; 0)

L(X, c, q, 0) = Sq(Zlc) logp(X,Zlc; 0)dZ - Sq(Zlc) logq(Zlc)dZ

= Sq(Zlc) logp(XIZ,c;0)dZ + Sq(Zlc) logp(Zlc)dZ

- Sq (Zlc) log q (Zlo) dZ

= Ez~q(Z|X,c;p) log p(X|Z,c;0) + Ez~q(Z|X,c;p) log p(Z|c)

- Ez~q(ZIX, cip) log q(ZIX, ci0)

= Ez-q(ZIX,c; d) log p(XIZ,c; 0) - KL (q(ZIX,c; 0) 11 p(Z|c))