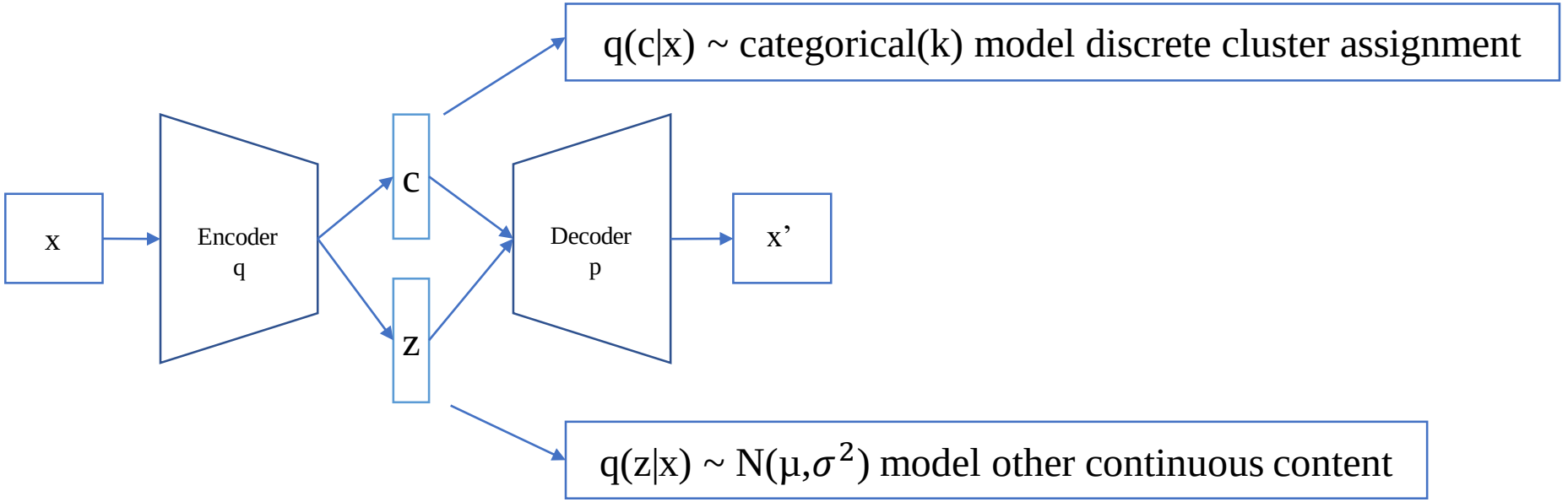

Loss: min $L_{VAE} - MI(x,c)$

$=L_{VAE} + H(X|C) - H(X)$

$= L_{VAE} - E_{c\sim p(c)}E_{x\sim p(x|c)}[logp(x|c)] + const$

$= E_q[logp(x|z,c)] - \beta D_{KL}[q(z,c)||p(z,c)] - E_{c\sim p(c)}E_{x\sim p(x|c)}[logp(x|c)] + const$

To enable back prop, q(c|x) is approximated by continuous Gumbel-softmax distribution

Combine VAE and Infomax, maximize the mutual information between data and cluster assignment