Kullback-Leibner Divergence

Kullback-Leibner Divergence is an information-theoretic measure that describes the difference between two probability distributions.

$$D_{KL}(q(\mathbf{z})||p(\mathbf{z})) = \mathbb{E}_{q(z)}[log(q(\mathbf{z}) - log(p(\mathbf{z})))] \tag{1}$$

Since we cannot obtain $p_{\hat{\theta}}(\mathbf{z}|\mathbf{x})$ in a direct solution we use the Kullback-Leibner Divergence to find another distribution that approximates $p_{\hat{\theta}}(\mathbf{z}|\mathbf{x})$.
Let $q(\mathbf{z}|\mathbf{x})$ define a pdf of $\mathbf{z}|\mathbf{x}$ and Θ a chosen parameter space. Then approximating $p_{\hat{\theta}}(\mathbf{z}|\mathbf{x})$ amounts to finding $\psi \in \Theta$ s.t. the Kullback-Leibner Divergence is minimal.

Let $\mathcal{F}_{q_{\mathbf{z}|\mathbf{x}}}$ be a parametric family of densities then $q^*_\psi(\mathbf{z}|\mathbf{x})$ is the best approximation to $p_{\hat{\theta}}(\mathbf{z}|\mathbf{x})$:

$$q^*_\psi(\mathbf{z}|\mathbf{x}) = \underset{q_\psi(\mathbf{z}|\mathbf{x}) \in \mathcal{F}_{q_{\mathbf{z}|\mathbf{x}}}}{\arg\min} D_{KL}(q_\psi(\mathbf{z}|\mathbf{x})||p_{\hat{\theta}}(\mathbf{z}|\mathbf{x})) \tag{2}$$