ELBO

变分推理的目标是近似**潜在变量 (latent variables) 在观测变量（observed variables）下的条件概率**。

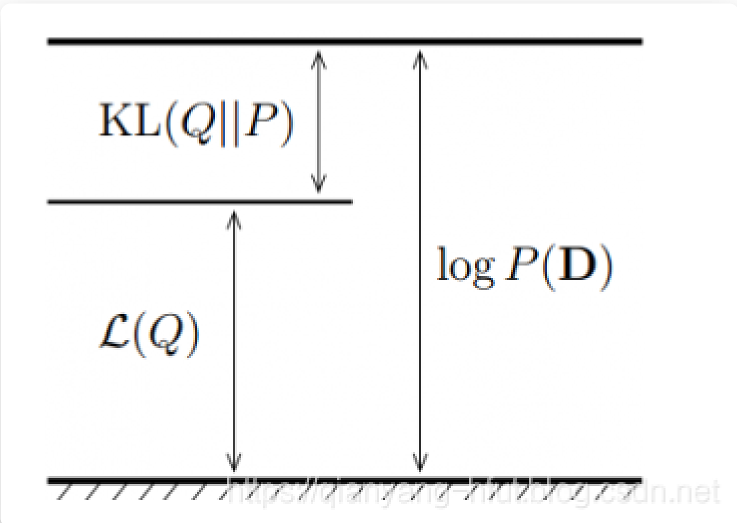
变分推断等价于最小化 KL 散度。

$$q^*(\mathbf{z}) = \underset{q(\mathbf{z})\in\mathcal{Q}}{\arg\min}\, \mathrm{KL}(q(\mathbf{z})\|p(\mathbf{z} \mid \mathbf{x}))$$

其中，$q(z)$ 为近似分布，$p(z|x)$为为 **所要求的的后验概率分布**。这里之所以对$p(z|x)$近似，因为其难以计算。KL 散度可以表示为：

$$KL[Q(x)\|P(x \mid D)] = \int dx \cdot Q(x) \ln \frac{Q(x)}{P(x \mid D)}$$

其中，$\ln P(D)$ 为 log 似然，L为log似然的下界。使得 KL 散度最小，相当于最大化L。如下为三者之间的关系：



$$
\begin{aligned}
D_{\mathrm{KL}}[q(\mathbf{Z}) \| p(\mathbf{Z} \mid \mathbf{X})] &= \int_q q(\mathbf{Z}) \log \frac{q(\mathbf{Z})}{p(\mathbf{Z} \mid \mathbf{X})} \\
&= \mathbb{E}_{q(\mathbf{Z})}\left[\log \frac{q(\mathbf{Z})}{p(\mathbf{Z} \mid \mathbf{X})}\right] \\
&= \underbrace{\mathbb{E}_{q(\mathbf{Z})}[\log q(\mathbf{Z})] - \mathbb{E}_{q(\mathbf{Z})}[\log p(\mathbf{Z}, \mathbf{X})]}_{-\,\mathrm{ELBO}(q)} + \log p(\mathbf{X}).
\end{aligned}
$$

$$
\mathrm{ELBO}(q) := \mathbb{E}_{q(\mathbf{Z})}[\log p(\mathbf{Z}, \mathbf{X})] - \mathbb{E}_{q(\mathbf{Z})}[\log q(\mathbf{Z})]
$$

$$
\log p(\mathbf{X}) = \mathrm{ELBO}(q) + D_{\mathrm{KL}}[q(\mathbf{Z}) \| p(\mathbf{Z} \mid \mathbf{X})]
$$

$$
\log p(\mathbf{X}) \geq \mathrm{ELBO}(q)
$$