Jensen's inequality

For a convex function h:

$$h(E[X]) \leq E[h(X)]$$

For a concave(logarithm) function h:

$$h(E[X]) \geq E[h(x)]$$ Lower bound

How Jensen inequality helps us: ELBO

$$\log E_{z\sim q_\phi(z)}\left[\frac{P_\theta(x,z)}{q_\phi(z)}\right] \geq E_{z\sim q_\phi(z)}\left[\log \frac{P_\theta(x,z)}{q_\phi(z)}\right]$$

$$\geq E_{z\sim q_\phi(z)} \log\left[\frac{P_\theta(x|z)P(z)}{q_\phi(z)}\right]$$

$$\geq E_{z\sim q_\phi(z)}\left[\log P_\theta(x|z) + \log P(z) - \log q_\phi(z)\right]$$

$$\geq E_{z\sim q_\phi(z)}\left[\log_\theta P(x|z)\right] - E_{z\sim q_\phi(z)}\left[\log q_\phi(z) - \log P(z)\right]$$

If we use amortized variational posterior $q_\phi(z|x)$ instead of $q_\phi(z)$, then

$$\log p(x) \geq E_{z\sim q_\phi(z|x)}\Big[\log \underbrace{P_\theta(x|z)}_{\text{Decoder}}\Big] - E_{z\sim q_\phi(z|x)}\Big[\log \underbrace{q_\phi(z|x)}_{\text{Encoder}} - \log P(z)\Big]$$

ELBO

$$\geq E_{z\sim q_\phi(z|x)}\left[\log P_\theta(x|z)\right] - KL\left[q_\phi(z|x) \,\|\, P(x)\right]$$
(1) (2)

$$\log P(x) \geq E_{z\sim q_\phi(z|x)}\left[\log P(x|z)\right] - E_{z\sim q_\phi(z|x)}\left[\log q_\phi(z|x) - \log P(z)\right].$$
(1) (2)

(1) term represents reconstruction, encouraging the reconstructions that maximize likelihood

(2) term minimizes the 'distance' between the density function $q_\phi(z|x)$ and prior $P(z)$.