ELBO Derivations (3 flavours)

1. From Jensen’s
2. From KL
3. From Bayes Rule Directly

1 Trick I: Multiply & divide

... this is expectation ...

$$\log p(\mathbf{x}) = \log \int p(\mathbf{x},\mathbf{z})dz = \log \int \frac{q(\mathbf{z})}{q(\mathbf{z})} p(\mathbf{x},\mathbf{z})dz = \log \mathbb{E}_{q(\mathbf{z})}\left[\frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z})}\right]$$

... after the inequality ... 2

$$\geq \mathbb{E}_{q(\mathbf{z})}\left[\log\left(\frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z})}\right)\right]$$

Chain Rule 3

$$\geq \mathbb{E}_{q(\mathbf{z})}\left[\log\left(\frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{q(\mathbf{z})}\right)\right]$$

... log properties ... 4

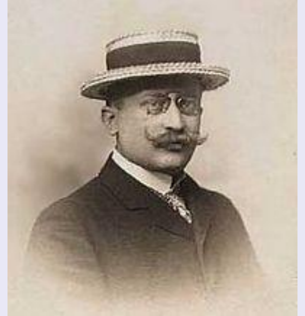
$$= \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{x}|\mathbf{z})] + \mathbb{E}_{q(\mathbf{z})}\left[\log\left(\frac{p(\mathbf{z})}{q(\mathbf{z})}\right)\right]$$

Def. of KL 5

Data-Likelihood Regularisation

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

The ELBO – Evidence Lower-Bound



Jensen's Inequality

$$\mathbb{E}[f(\mathbf{x})] \leq f(\mathbb{E}[x])$$

... for <u>concave</u> fncs ...

... apply the inequality ...

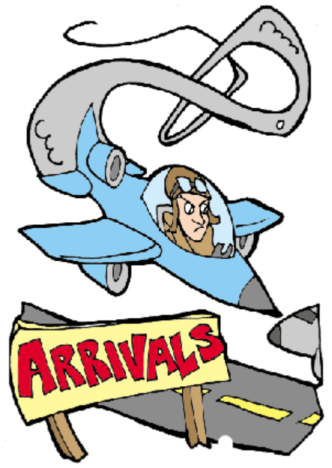
$$f(\mathbb{E}[x]) = \log\left(\mathbb{E}_{q(\mathbf{z})}[\cdot]\right) \geq \mathbb{E}[f(\mathbf{x})] = \mathbb{E}_{q(\mathbf{z})}[\log(\cdot)]$$

Start from the goal (close-dist. to the posterior)

... our approximation dist. ...

$$\mathrm{KL}(q(\mathbf{z})||p(\mathbf{z}|\mathbf{x})) = \int q(\mathbf{z}) \log \left[\frac{q(\mathbf{z})}{p(\mathbf{z}|\mathbf{x})}\right] d\mathbf{z}$$

... the posterior (our approx. goal) ...



... arrive again at the Evidence Lower-Bound (ELBO).

$$\mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{x}|\mathbf{z})] - \mathrm{KL}(q(\mathbf{z})||p(\mathbf{z}))$$

1 ● Log props. ● Multiply dist. ● Integral props.

Rule of Conditional Probabilities

$$\int q(\mathbf{z}) \log \left[\frac{q(\mathbf{z})}{p(\mathbf{z}|\mathbf{x})}\right] d\mathbf{z} = \int q(\mathbf{z}) \log q(\mathbf{z}) d\mathbf{z} - \int q(\mathbf{z}) \log p(\mathbf{z}|\mathbf{x}) d\mathbf{z} = \int q(\mathbf{z}) \log q(\mathbf{z}) d\mathbf{z} - \int q(\mathbf{z}) \log \frac{p(\mathbf{x},\mathbf{z})}{p(\mathbf{x})} d\mathbf{z}$$

$$= \int q(\mathbf{z})(\log q(\mathbf{z}) - \log p(\mathbf{x},\mathbf{z})) d\mathbf{z} + \int q(\mathbf{z}) \log p(\mathbf{x}) d\mathbf{z}$$

2 ● Log props. ● Multiply dist. ● Integral props.

3 Those integrals are expectations

$$= \mathbb{E}_{q(\mathbf{z})}[\log q(\mathbf{z}) - \log p(\mathbf{x},\mathbf{z})] + \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{x})]$$

4 ● Independent RV

$$\mathrm{KL}(q(\mathbf{z})||p(\mathbf{z}|\mathbf{x})) = \mathbb{E}_{q(\mathbf{z})}[\log q(\mathbf{z}) - \log p(\mathbf{x},\mathbf{z})] + \log p(\mathbf{x})$$

Saw it before!

(same as 3,4,5 of previous slide)

5 Kullback-Leibler is always positive

$$\mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{x},\mathbf{z}) - \log q(\mathbf{z})] = \log p(\mathbf{x}) - \mathrm{KL}(q(\mathbf{z})||p(\mathbf{z}|\mathbf{x})) \implies \log p(\mathbf{x}) \geq \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{x},\mathbf{z}) - \log q(\mathbf{z})]$$

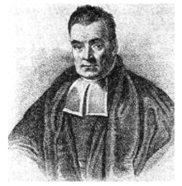
6 Prob. Chain rule

$$= \mathbb{E}_{q(\mathbf{z})}\left[\log\left(\frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z})}\right)\right] = \mathbb{E}_{q(\mathbf{z})}\left[\log\left(\frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{q(\mathbf{z})}\right)\right]$$

Arrivals: Evidence-Lower Bound (ELBO)

7 ● Log props. ● Def of KL

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



1 Start from Bayes Rule

Posterior, Likelihood, Prior, Normaliser

$$p(\mathbf{z}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{x})}$$

2 Cross-Multiply

Prob. Rule

$$\ln p(\mathbf{x}) = \ln\left[\frac{p(\mathbf{x}|\mathbf{z})p(\mathbf{z})}{p(\mathbf{z}|\mathbf{x})}\right] \qquad \ln p(\mathbf{x}) = \ln p(\mathbf{x},\mathbf{z}) - \ln p(\mathbf{z}|\mathbf{x})$$

Ln props.

3 Trick I: Add & Subtract

$$0 = -\ln q(\mathbf{z}) + \ln q(\mathbf{z})$$

... unchanged ...

$$\ln p(\mathbf{x}) = \ln p(\mathbf{x},\mathbf{z}) - \ln p(\mathbf{z}|\mathbf{x}) - \ln q(\mathbf{z}) + \ln q(\mathbf{z})$$

... this trick introduced the **variational distribution** ...

$$\ln p(\mathbf{x}) = \ln p(\mathbf{x},\mathbf{z}) - \ln q(\mathbf{z}) - \ln \frac{p(\mathbf{z}|\mathbf{x})}{q(\mathbf{z})}$$

... re-arrange & log props ...

The Ln Bound

$$\ln x \leq x - 1 \rightarrow -\ln x \geq 1 - x$$

apply

4

$$\ln p(\mathbf{x}) \geq \ln p(\mathbf{x},\mathbf{z}) - \ln q(\mathbf{z}) + 1 - \frac{p(\mathbf{z}|\mathbf{x})}{q(\mathbf{z})}$$

... true for all z ...

5 Trick II: Expectation

● Log props. ● Multiply dist. ● Cancel numerator & denominator

... this is just 1 (pdf int.) ...

$$\ln p(\mathbf{x}) \geq \int q(\mathbf{z})\left(\ln p(\mathbf{x},\mathbf{z}) - \ln q(\mathbf{z}) + 1 - \frac{p(\mathbf{z}|\mathbf{x})}{q(\mathbf{z})}\right)d\mathbf{z} = \int q(\mathbf{z}) \ln \frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z})} d\mathbf{z} + 1 - \int p(\mathbf{z}|\mathbf{x}) d\mathbf{z}$$

... true for all z -> then also for any sample from any distribution ...

ELBO $= \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{x}|\mathbf{z})] - \mathrm{KL}(q(\mathbf{z})||p(\mathbf{z}))$

Saw it before!

● Chain Rule ● Log props. ● Def of KL