$$\ln q_j^*(z_j) = \mathbb{E}_{i \neq j}[\ln p(X, z)] + \text{const}$$

lacks normalization (const)

- only yields $q_j^*$ up to a multiplicative factor
- one can normalize this distribution by either
- (usually)
  - inspection (will be become clear)
  - or by explicit normalization

$$q_j^*(z_j) = \frac{\exp(\mathbb{E}_{i \neq j}[\ln p(x, z)])}{\int \exp(\mathbb{E}_{i \neq j}[\ln p(x, z)])\, dz_j}$$

- Set of the eqn's for all $q_i$, $i = 1..M$ is a set of "consistencies" conditions for the max. They are not an explicit sol'n, not in general closed form, and have to cycled through until numerical convergence.

State without proof Convergence of these interdependent updates is guaranteed because the objective is convex.

## Teaching Example

Variational approximation to a full covariance Gaussian (2D)

Remember, in general $\vec{z} \in \mathbb{R}^2$ $\vec{z} \sim N(\mu, \Sigma)$

$$p(\vec{z}) \neq p(z_1)\, p(z_2) \quad \text{unless} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and $z_1 \perp\!\!\!\perp z_2$

Goal: find the independent Gaussian dist (diagonal Gaussian) that best approximates $p(\vec{z})$

Factorization $q(z_1)\, q(z_2)$

covariance

Let's have

$$p(\vec{z}) = \mathcal{N}(\vec{z} \mid \mu, \Lambda^{-1})$$ "complicated"

$$\mu = [\mu_1, \mu_2]^T \qquad \underset{\uparrow \text{ precision}}{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$$ where $\lambda_{12} = \lambda_{21}$

Suppose we want to find the best approximating dist'n of the form

$$q(\vec{z}) = q_1(z_1)\, q_2(z_2)$$

Note: a bit of a weird example, there is no data
- marginalizing over all possible draws from $p(\vec{z})$, $z$ is itself the latent quantity.

Start with $q_1(z_1)$, apply the general recipe to find $q_1^*(z_1)$

$$\ln q_1^*(z_1) = \mathbb{E}_{q_2(z_2)}\left[\ln p(\vec{z})\right] + \text{const}$$

↑ all terms not a function of $q_1$

Remember $\ln p(\vec{z}) \propto -\frac{1}{2}(z-\mu)^T \Lambda (z-\mu)$

$$= -\frac{1}{2}\left( [z_1-\mu_1 \quad z_2-\mu_2] \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \begin{bmatrix} z_1-\mu_1 \\ z_2-\mu_2 \end{bmatrix} \right)$$

$$-\frac{1}{2}\left( \lambda_{11}(z_1-\mu_1)^2 + \lambda_{12}(z_1-\mu_1)(z_2-\mu_2) + \lambda_{22}(z_2-\mu_2)^2 \right)$$

$$\ln q_1^*(z_1) = \mathbb{E}_{q_2(z_2)}\left[\ln p(\vec{z})\right] + \text{const}$$

$$= \mathbb{E}_{q_2(z_2)}\left[-\frac{1}{2}\lambda_{11}(z_1-\mu_1)^2 - \lambda_{12}(z_1-\mu_1)(z_2-\mu_2) - \frac{1}{2}\lambda_{22}(z_2-\mu_2)\right]$$

const don't involve $z_1$

$$= -\frac{1}{2}\lambda_{11}(z_1-\mu_1)^2 - \mathbb{E}_{q_2(z_2)}\left[\lambda_{12}(z_1 z_2 - z_1\mu_2 - \mu_1 z_2 + \mu_1\mu_2)\right] + \text{const}$$

const w.r.t. $z_1$

$$= -\frac{1}{2}\lambda_{11}(z_1-\mu_1)^2 - \lambda_{12}\, z_1\left(\mathbb{E}_{q_2(z_2)}[z_2] - \mu_2\right) + \text{const}$$

by inspection this is a Gaussian distribution

result is a quadratic from in $z_1$, $\Rightarrow z_1$ is Gaussian distributed, to find the params of said Gaussian we need to "complete the square"

Complete the Square

. Recognize that $z_1 \sim N(m_1, S_1)$ because quad. form
$\quad ? \quad ?$

$$\ln p(z_1 \mid m_1, S_1) \propto -\frac{1}{2S_1}(z_1-m_1)^2$$

$$\propto -\frac{1}{2S_1}(z_1^2 - 2z_1 m_1 + m_1^2)$$

$$\propto -\frac{z_1^2}{2S_1} + \frac{z_1 m_1}{S_1} + \text{const terms in } z_1$$

first ~~term~~ identifies the variance

second term can be used to solve for the mean

$$\ln q_1^*(z_1) = -\frac{1}{2}\lambda_{11}(z_1-\mu_1)^2 - \lambda_{12} z_1 \left(\mathbb{E}_{q_2}[z_2] - \mu_2\right) + \text{const}$$

expanding & grouping

$$= -\frac{1}{2}\lambda_{11}\left(z_1^2 - 2\mu_1 z_1 + \overset{\text{const in } z_1}{\cancel{\mu_1^2}}\right) - \lambda_{12} z_1 \left(\mathbb{E}_{q_2}[z_2] - \mu_2\right) + \text{const}$$

$$= \underbrace{-\frac{1}{2}\lambda_{11} z_1^2}_{S_1 = \lambda_{11}^{-1}} + z_1\left(\lambda_{11}\mu_1 - \lambda_{12}\left(\mathbb{E}_{q_2}[z_2] - \mu_2\right)\right)$$

$S_1 = \lambda_{11}^{-1}$ ← gives us the variance of $q_1^*(z_1)$

can solve for mean, $m_1$

$$\frac{z_1 m_1}{S_1} = z_1\left(\lambda_{11}\mu_1 - \lambda_{12}\left(\mathbb{E}_{q_2}[z_2] - \mu_2\right)\right)$$

$$m_1 = \mu_1 - \lambda_{11}^{-1}\lambda_{12}\left(\mathbb{E}_{q_2}(z_2) - \mu_2\right)$$

So

$$q_1^*(z_1) = \mathcal{N}(z_1 \mid m_1, S_1)$$

$$= \mathcal{N}\left(z_1 \mid \mu_1 - \lambda_{11}^{-1}\lambda_{12}\left(\mathbb{E}_{q_2}(z_2) - \mu_2\right), \lambda_{11}^{-1}\right)$$

Problem completely symmetric

$$q_2^*(z_2) = \mathcal{N}\left(z_2 \mid \mu_2 - \lambda_{22}^{-1}\lambda_{12}\left(\mathbb{E}_{q_1}(z_1) - \mu_1\right)\right)$$

So what is $\mathbb{E}_{q_2}(z_2)$? $= m_2$

$$\mathbb{E}_{q_1}(z_1) = m_1$$

Sol'ns are <u>coupled</u>! But iterative procedure to find max is obvious. Example code for this example is online. ◻