A Gaussian Reparametrisation

Any continuous variable can be transformed into another one by an invertible and differentiable transformation h. While we can compute the density of the transformed variable at a given point, we can generally not sample from that density. Fortunately, Gaussian variables can be transformed into standard Gaussian variables. The standard Gaussian distribution is easy to sample from. Let $z \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and recall that in the inference network μ and σ are functions of the variational parameters λ which we seek to optimise. We can standardise z by subtracting its mean and dividing by the standard deviation,

$$\epsilon = h(z, \mu, \sigma) = \frac{z - \mu}{\sigma} \,. \tag{16}$$

We write the standard Gaussian density of $\epsilon \sim \mathcal{N}\left(0, \mathrm{I}\right)$ as $\phi(\epsilon)$.

Differentiating the reconstruction term of the ELBO with respect to the variational parameters λ poses a challenge. If we differentiate the expectation we end up with an expression that is itself not an expectation and can thus not be approximated with MC methods (see Equation (11) of the main text). In order to enable MC estimation we replace the measure of the expectation with its transformed version.

$$\begin{aligned} q(z|\mu, \sigma^2) &= \phi(h(z, \mu, \sigma)) \times \left|\frac{d}{dz} h(z, \mu, \sigma)\right| \\ &= \phi(\epsilon) \times \left|\frac{d\epsilon}{dz}\right| \end{aligned} \tag{17}$$

The Jacobian term corrects for the change in volume introduced by the transformation.

Taking the derivative of the ELBO with respect to the variational parameters λ is now easy as we can push the derivative operator inside the expectation to obtain stochastic gradient estimates.

$$\frac{\partial}{\partial \lambda} \mathbb{E}_{q(z|\mu, \sigma^2)} \left[\log p(y|x, z)\right] = \tag{18a}$$

$$\frac{\partial}{\partial \lambda} \int q(z|\mu, \sigma^2) \log p(y|x, z) \mathrm{d}z = \tag{18b}$$

$$\frac{\partial}{\partial \lambda} \int \phi(\epsilon) \times \left|\frac{d\epsilon}{dz}\right| \log p(y|x, z) \mathrm{d}\epsilon \left|\frac{dz}{d\epsilon}\right| = \tag{18c}$$

$$\int \phi(\epsilon) \times \frac{\partial}{\partial \lambda} \log p(y|x, z) \mathrm{d}\epsilon = \tag{18d}$$

$$\mathbb{E}_{\phi(\epsilon)} \left[\frac{\partial}{\partial \lambda} \log p(y|x, \underbrace{h^{-1}(\epsilon, \mu, \sigma)}_{z}) \right] \tag{18e}$$

Notice that the inverse transformation h^{-1} is exactly the one described in Equation (5) of the main text. In line (18c) we have used the substitution rule. As a result the Jacobian terms cancel, leaving an integral whose measure is standard normal.