Variational approximations Variational methods define a lower bound

$$\mathcal{L}(\boldsymbol{x};\boldsymbol{\theta}) \leq \log p_{\text{model}}(\boldsymbol{x};\boldsymbol{\theta}). \tag{7}$$

A learning algorithm that maximizes $\mathcal{L}$ is guaranteed to obtain at least as high a value of the log-likelihood as it does of $\mathcal{L}$. For many families of models, it is possible to define an $\mathcal{L}$ that is computationally tractable even when the log-likelihood is not. Currently, the most popular approach to variational learning in deep generative models is the **variational autoencoder** (Kingma, 2013; Rezende *et al.*, 2014) or VAE. Variational autoencoders are one of the three approaches to deep generative modeling that are the most popular as of this writing, along with FVBNs and GANs. The main drawback of variational methods is that, when too weak of an approximate posterior distribution or too weak of a prior distribution is used, [1] even with a perfect optimization algorithm and infinite training data, the gap between $\mathcal{L}$ and the true likelihood can result in p_{model} learning something other than the true p_{data}. GANs were designed to be unbiased, in the sense that with a large enough model and infinite data, the Nash equilibrium for a GAN game corresponds to recovering p_{data} exactly. In practice, variational methods often obtain very good likelihood, but are regarded as producing lower quality samples. There is not a good method of quantitatively measuring sample quality, so this is a subjective opinion, not an empirical fact. See figure 11 for an example of some samples drawn from a VAE. While it is difficult to point to a single aspect of GAN design and say that it results in better sample quality, GANs are generally regarded as producing better samples. Compared to FVBNs, VAEs are regarded as more difficult to optimize, but GANs are not an improvement in this respect. For more information about variational approximations, see chapter 19 of Goodfellow *et al.* (2016).

[1] Empirically, VAEs with highly flexible priors or highly flexible approximate posteriors can obtain values of $\mathcal{L}$ that are near their own log-likelihood (Kingma *et al.*, 2016; Chen *et al.*, 2016b). Of course, this is testing the gap between the objective and the bound at the maximum of the bound; it would be better, but not feasible, to test the gap at the maximum of the objective. VAEs obtain likelihoods that are competitive with other methods, suggesting that they are also near the maximum of the objective. In personal conversation, L. Dinh and D. Kingma have conjectured that a family of models (Dinh *et al.*, 2014; Rezende and Mohamed, 2015; Kingma *et al.*, 2016; Dinh *et al.*, 2016) usable as VAE priors or approximate posteriors are universal approximators. If this could be proven, it would establish VAEs as being asymptotically consistent.