S1 Appendix. Blackbox Variational Inference

We explain the derivation for Blackbox Variational Inference (BBVI, [1], [2]). In the following discussion, we use $x$ to refer to observed data (i.e. regional infection statistics), and $z$ to refer to the set of all latent variables (such as the initial community exposure rates $\rho_{1:C}$ and the exposure probabilities $\beta^{E}_{1:t_N}, \beta^{I}_{1:t_N}$). We use Stochastic Variational Inference (SVI) to approximate the intractable posterior $p(z|x)$ by optimizing the parameters of a variational distribution $q(z; \phi)$ (also denoted $q_\phi$) from which we are able to directly draw samples.

**ELBO Objective.** A common approach to optimizing the variational distribution is to find parameters $\phi^*$ that minimize the exclusive Kullback-Leibler (KL) divergence between the variational approximation and the posterior:

$$\phi^* = \arg\min_\phi \text{KL}(q_\phi(z) \mid \mid p(z \mid x)),$$

Minimizing this exclusive KL divergence is equivalent to maximizing a lower bound on the log marginal likelihood of the data. Since the marginal likelihood $p(x)$ is also referred to as the “evidence”, this bound is called the Evidence Lower BOund (ELBO),

$$\mathcal{L}_\phi = \mathbb{E}_{z \sim q_\phi} \left[ \log \frac{p(x, z)}{q_\phi(z)} \right] = \log p(x) - \text{KL}(q_\phi(z) \mid \mid p(z \mid x)) \leq \log p(x). \quad (4)$$

**Score Function ELBO Gradient.** Variational Autoencoders [3, 4] and related methods optimize the ELBO by computing reparametrized gradient estimates, which require the generative model $p(x, z)$ to be differentiable with respect to the latent variables $z$. These reparametrized gradient estimators have the advantage of low variance (meaning that relatively little sampling and computation is required per gradient step), but impose limits on the generative models being studied. Specifically, models that incorporate discrete variables and control flow will not be differentiable, and it may be infeasible or undesirable to find a continuous approximation to make such a model differentiable. We therefore choose to maximize the ELBO using a so-called “score-function” gradient estimator that does not require reparameterization. This approach is generally referred to as Blackbox Variational Inference (BBVI) [1, 2] We repeat the derivation of this gradient estimator here for convenience,
While it is not guaranteed to reduce the variance and introduces a small bias for finite sample sizes, it is a standard choice and tends to work well in practice (see [10] for a review).
References


