# A Weinberg benchmark

## A.1 Simulation

For this benchmark inference task, we consider a simplified simulator from particle physics for electron–positron collisions resulting in muon–antimuon pairs ($e^+e^- \to \mu^+\mu^-$). The simulator approximates the distribution of observed measurements $\mathbf{x} = \cos(A) \in [-1, 1]$, where $A$ is the polar angle of the outgoing muon with respect to the originally incoming electron. Neglecting measurement uncertainty induced from the particle detectors, this random variable is approximately distributed as

$$p(\mathbf{x}|E^{\text{beam}}, G^f) = \frac{1}{Z}\left[(1+\mathbf{x}^2) + c(E^{\text{beam}}, G^f)\mathbf{x}\right]$$

where $Z$ is a known normalization constant and $c$ is an asymmetry coefficient function. Due to the linear term in the expression, the density $p(\mathbf{x}|E^{\text{beam}}, G^f)$ exhibits a so-called *forward-backward* asymmetry. Its size depends on the values of the parameters $E^{\text{beam}}$ (the beam energy) and $G^f$ (the Fermi constant) through the coefficient function $c$.

A typical physics simulator for this process includes a more precise treatment of the quantum mechanical $e^+e^- \to \mu^+\mu^-$ scattering using Pythia or MadGraph (Alwall et al., 2011), ionization of matter in the detector due to the passage of the out-going $\mu^+\mu^-$ particles using GEANT4 (Agostinelli et al., 2003), electronic noise and other details of the sensors that measure the ionization signal, and the deterministic algorithms that estimate the polar angle $A$ based on the sensor readouts. The simulation of this process is highly non-trivial as is the space of latent variables $\mathcal{Z}$.

## A.2 Results

A prominent issue with the Weinberg benchmark is the presence of a nearly degenerate direction for the likelihood in the model parameter space. This leads to a number of solutions that provide good fits to the observed data. Since Figure 4 evaluates $||\boldsymbol{\theta}^* - \hat{\boldsymbol{\theta}}||_2^2$, the presence of this broad minima significantly influences the result. To show that AVO, SMC-ABC, and BOLFI do find solutions that describe the data well, we sample $\mathbf{x} \sim p(\mathbf{x}|\hat{\boldsymbol{\theta}})$ (inferred) and compare against $p_r(\mathbf{x})$ (observed) for several $\boldsymbol{\theta}_i^*$, as shown in Figure 5. These plots demonstrate that for this benchmark, there exist many equivalent solutions that induce the observed data, even if they may be quite distant in parameter space.

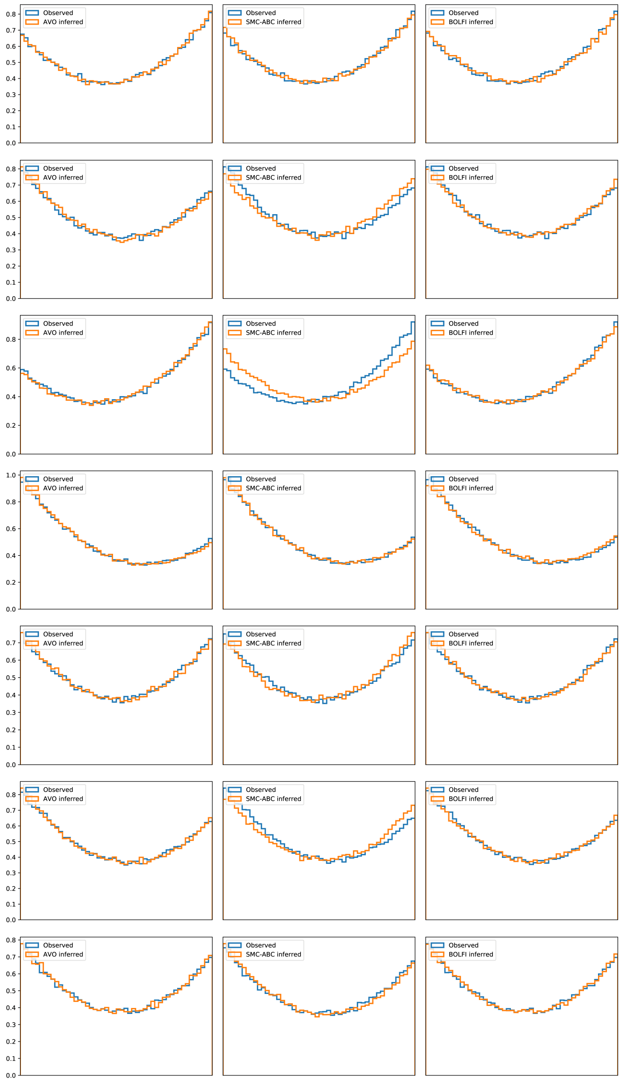

Figure 5: (*Left*) AVO. (*Center*) SMC-ABC. (*Right*) BOLFI. Despite the apparent poor performance of AVO, SMC-ABC and BOLFI in Figure 4, all methods approximate the observed data distribution $p_r(\mathbf{x})$ for different $\boldsymbol{\theta}_i^*$ (rows). This discrepancy is attributed to multiple minima in the Weinberg benchmark.