Inference

We built an inference framework to be able to calibrate simulated epidemic trajectories to observed epidemic data. Calibrated parameters of our modeling framework, which are all potentially location-specific, include epidemic seeding dates and amounts, values of the basic reproduction number, and the effectiveness of different types of NPIs. Inference on model parameters for each spatial location are drawn jointly given the spatial coupling of COVID-19 transmission dynamics.

Fit Parameter

Fit parameters are user specified, and can include NPI effectiveness and the timing and number of introduced cases that seed the epidemic.

Pseudo-Likelihood

We enable the user to specify the likelihood function in terms of the epidemiological data used for calibration (e.g., number of incident confirmed cases or deaths) and the probability distributions associated with those calibration data. We consider the likelihood of the set of parameters for a location i, Θ_i, given some set of Y_i to be equal to likelihood of that simulation given the data:

$$L(\Theta_i|Y_i) = L(Z_i|Y_i)$$

Where Z is a full epidemic simulation (jointly generated for all locations) generated from a parameter set Θ. Note that since epidemic simulations are not stochastic, this value will not be deterministically determined for any parameter set.

Given a location i, a set of J data variables, and T time units (e.g., days, weeks), the likelihood of a simulate epidemic for a location Z_i given the observed data Y_i at that location is :

$$L(Z_i; Y_i) = \prod_{j=1}^{J} \prod_{t=1}^{T} p_j(Y_{i,j,t}|Z_{i,j,t})$$

where p_j is the user-specified probability distribution for data variable j. These location-specific likelihoods to compute the full simulation likelihood:

$$L(\theta; Y) = L(Z; Y) = \prod_{i=1}^{N} L_i(Z_i; Y_i)$$

where N is the number of modeled locations. The framework allows the user to specify the time unit for a given calibration data; for instance one may calibrate to weekly incident deaths, biweekly confirmed cases, or both.

Fitting Algorithm

We attempt to obtain a posterior distribution over both parameters Θ and epidemic trajectories Y using a modified block MCMC algorithm designed to take advantage of highly parallel computing resources. In its essence, this algorithm runs a larger number of relatively

short MCMC chains, and uses the final accepted value from each chain to approximate the posterior distribution.

Each MCMC chain explores both parameter space and the space of epidemic trajectories in a way designed to efficiently explore parameter space in a way that aggressively maximizes individual local likelihoods, while preserving globally consistent epidemic trajectories. Each MCMC iteration k proceeds as follows:

1) generate proposed parameter set Θ_{prop}.
2) stochastically generate epidemic trajectory Z_{prop} using parameter set Θ_{prop}
3) **if (**$\Pr(Y|Z_{prop}/Pr(Y|Z_{(k-1)}) > runif(0,1)$**)**
   set $Z_{(k)} = Z_{prop}$
   set $\Theta^*_{(k)} = \Theta_{prop}$
   else
   set $Z_{(k)} = Z_{(k-1)}$
   set $\Theta^*_{(k)} = \Theta^*_{(k-1)}$
   endif
4) **for** (i **in** 1...N)
   if ($\Pr(Y_i|Z_{i,prop})/(Y_i|Zi,(k-1)) > runif(0,1)$**)**
   set $\Theta_{i,(k)} = \Theta i, prop$
   else
   set $\Theta_{i,(k)} = \Theta_{i,(k-1)}$
   endif
   endfor

This process is repeated a fixed number K times and for each of M slots, with $Z_{(k)}$ and $\Theta^*_(K)$ being the final returned value for each slot. This final set of values is considered to represent both the posterior fit distribution of parameters and the posterior predictive distribution over epidemic trajectories.

Proposal distribution (i.e., perturbations) for parameters can be specified in the config file.