Incomplete data in the COVID-19 pandemic

Michael Baron

[1] American University, Washington, USA

Abstract:
Officially reported COVID-19 daily counts of infected, recovered, and perished people are substantially underestimated for a number of reasons. Only a portion of infected individuals receives professional coronavirus testing. Among those who have been tested, confirmed, and then recovered, only a portion of recoveries are reported. The proportion of unobserved and under-observed counts varies by territory and changes in time, due to different and changing diagnostics and reporting standards.

To remedy the incomplete data, we develop a stochastic model that includes untested individuals and unobserved COVID-19 recoveries and casualties. It generalizes the classical Susceptible-Infected-Removed (SIR) epidemic model, extending it with additional compartments. Its main parameters are the rates of infection, testing, vaccination, recovery, mortality, and reporting, which may vary in time. The proposed Bayesian algorithm uses observed counts to estimate the model parameters and unobserved counts dynamically, updating the estimates with new data on daily basis and resulting in a mutually consistent set of counts.

Keywords:
Epidemic modeling, SIR, Bayesian estimation, Markov chain, transition rate