Bayesian SEIR Model

Author: Dan Sheldon
Date: March 31, 2020

Underlying Disease Model

We use a a standard SEIR model with a time-varying contact rate:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta(t) \cdot S \frac{I}{N} \\
\frac{dE}{dt} &= \beta(t) \cdot S \frac{I}{N} - \sigma E \\
\frac{dI}{dt} &= \sigma E - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}$$

One additional variable $C(t)$ is added to track cumulative number of infections for the purposes of the observation model:

$$\frac{dC}{dt} = \sigma E$$

The nature of $\beta(t)$ will be discussed below --- it will change stochastically on a daily basis, but is constant for the duration of each day.

Stochastic Model

Parameters and Intial State Variables

$$\begin{aligned}
I_0 &\sim \text{Unif}(0, 0.001N) \\
E_0 &\sim \text{Unif}(0, 0.001N) \\
\sigma &\sim \text{Gamma}(5, 5\hat{d}_E) \\
\gamma &\sim \text{Gamma}(5, 5\hat{d}_I) \\
\beta_0 &\sim \text{Gamma}(1, \hat{d}_I/\hat{R}) \\
\rho &\sim \text{Beta}(\kappa\hat{\rho}, \kappa(1-\hat{\rho}))
\end{aligned}$$

Justification and values for user-selected parameters:

- I_0 and E_0 are the initial numbers of infected and exposed. The priors are self-explanatory.
- σ is the rate for leaving the exposed compartment; i.e., $1/\sigma$ is the expected duration in the exposed compartment.

The prior satisfies $\mathbb{E}[\sigma] = 1/\hat{d}_E$, where $\hat{d}_E$ is an initial guess of the duration in the exposed compartment. Currently $\hat{d}_E = 4.5$ based on Lauer et al., who estimated the median incubation period to be 5.1 days --- shortened by 0.6 days for possible infectiousness prior to developing symptoms

- γ is the rate for leaving the infectious compartment; i.e., $1/\gamma$ is the expected duration in the infectious compartment. The prior satisfies $\mathbb{E}[\gamma] = 1/\hat{d}_I$, where $\hat{d}_I$ is an initial guess for the duration in the infectious compartment. The current setting is $\hat{d}_I = 1.5$ to model the likely isolation of individuals after symptom onset. Preliminary experiments show that the model has almost no ability to estimate this parameter from available data --- it is too confounded with σ and β. Initial experiments also showed that $\hat{d}_I$ needed to be pretty small to be able to fit observed data while keeping realistic estimates of incubation time and R_0 . It's probably important to do sensitivity analyses of forecasts on the setting of this prior.
- β is the contact rate. In the SEIR model, it is known that $R_0 = \beta/\gamma$, so we set our prior to have mean $\mathbb{E}[\beta] = \hat{R}/\hat{d}_I$ where $\hat{R} = 3.5$ is an initial guess for R_0 and $\hat{d}_I = 1.5$, as described above.
- ρ is the detection rate. This prior satisfies $\mathbb{E}[\rho] = \hat{\rho}$ where $\hat{\rho} = 0.3$ is an initial guess for the detection rate. The parameter κ is a concentration parameter, currently set to $\kappa = 50$

Philosophy for shape and concentration parameters:

- Case count data effectively places one strong constraint on (β, γ, σ) , so we probably need informative priors on two of these.
- Our better outside evidence is probably on incubation period and infectious period, as opposed to contact rate, we we set stronger priors (shape = 5) on σ and γ, and a weaker prior (shape=1) on β_0 .
- The concentration parameter κ is set somewhat arbitrarily.

Process model

The process model proceeds in discrete time steps $k = 1, 2, 3, \ldots$.

The state variables are initialized as:

$$X(0) = \Big(S(0), E(0), I(0), R(0), C(0)\Big) = \Big(N - I_0 - E_0, E_0, I_0, 0, I_0\Big)$$

The updates are

$$\begin{aligned} \beta_{k+1} &= \beta_k \times \exp(\epsilon_k), \qquad \epsilon_k \sim \mathcal{N}(0, \tau) \\ X(k+1) &= \texttt{ode\_solve}\Big(X(k),\ dX/dt,\ \beta_k\Big) \end{aligned}$$

The contact rate β_k undergoes an exponentiated Gaussian random walk starting from β_0 (defined above). The function `ode_solve` finds the state vector at time $k + 1$ by simulating the ODE for one time step with *constant* contact rate β_k . (This could be replaced by something like Euler's approximation, Runge-Kutta, or even stochastic dynamics).

Observation model

Observations are made on confirmed cases. Let $y(t)$ be the cumulative number of confirmed cases at time t. The model is:

$$y(t) \sim \text{BinomApprox}\Big(C(t), \rho, \lambda\Big)$$

The distribution BinomApprox is a continuous approximation of a binomial distribution that can also model overdispersion via the parameter λ. The definition is

$$\text{BinomApprox}(n, p, \lambda) = n \cdot \text{Beta}\Big(\lambda p, \lambda(1-p)\Big)$$

With $\lambda = n$, we get the binomial. With $\lambda < n$, the distribution is overdispersed. We use $\lambda = 10$, so the variance is roughly the same as a Binomial distribution with $n = 10$. (TODO: show a picture of this.)