Bayesian SEIRD Model

Authors: Dan Sheldon and Casey Gibson

Date: April 27, 2020

Underlying Disease Model

We use an SEIR model augmented with two additional compartments: "D" for death, and "H" for "hospitalized-and-will-die". Note that H does not model all hospitalizations, only those that eventually lead to death.

$$\frac{dS}{dt} = -\beta(t)\,\frac{SI}{N}$$

$$\frac{dE}{dt} = \beta(t) \cdot \frac{SI}{N} - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = (1 - \rho)\gamma I$$

$$\frac{dH}{dt} = \rho\gamma I - \lambda H$$

$$\frac{dD}{dt} = \lambda H$$

The parameters are:

- $\beta(t)$: (time-varying) contact rate
- σ: rate of transition from E to I
- γ: rate of transition from I to (R/H)
- ρ: fatality rate (i.e., probability of transitioning from I to H instead of I to R)
- λ: rate of transition from H to D (inverse of expected number of days in H compartment before death)

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

$$I_0 \sim \text{Unif}(0, 0.02N)$$
$$E_0 \sim \text{Unif}(0, 0.02N)$$
$$H_0 \sim \text{Unif}(0, 0.02N)$$
$$D_0 \sim \text{Unif}(0, 0.02N)$$
$$\sigma \sim \text{Gamma}(5, 5\hat{d}_E)$$
$$\gamma \sim \text{Gamma}(7, 7\hat{d}_I)$$
$$\beta_0 \sim \text{Gamma}(1, \hat{d}_I/\hat{R})$$
$$\lambda \sim \text{Gamma}(10, 100)$$
$$\rho \sim \text{Beta}(10, 90)$$
$$p \sim \text{Beta}(15, 35)$$
$$p_d \sim \text{Beta}(90, 10)$$

Justification and values for user-selected parameters:

- I_0, E_0, H_0, and D_0 are the initial numbers of infectious, exposed, hospitalized-and-will-die, and dead. 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 $\hat{d}_E = 4.0$ based on published estimates (shortened slightly to account 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_0 is the initial contact rate. In the SEIR model, it is known that $R_0 = \beta/\gamma$, so we set our prior to have mean $\mathbb{E}[\beta_0] = \hat{R}/\hat{d}_I$ where $\hat{R} = 3.5$ is an initial guess for R_0 and $\hat{d}_I = 1.5$, as described above.
- λ is the inverse of the expected number of days in the H compartment; it satisfies $\mathbb{E}[\lambda] = 0.1$ with shape 10 (i.e. roughly 10 days in the H compartment)
- ρ is the fatality rate: it satisfies $\mathbb{E}[\rho] = 0.1$ with concentration of 100
- p is the detection probability: it satisfies $\mathbb{E}[p] = 0.3$ with concentration 50
- p_d is the detection probability for deaths: it satisfies $\mathbb{E}[p_d] = 0.9$ with concentration 100

Informative priors are useful on at least some of the internal parameters, since observed cases alone essentially place only one constraint on (σ, γ, β) --- we put more informative priors on (σ, γ) and allow the model more freedom to estimate β. We have made some but not extensive efforts to keep priors on other parameters less informative. Certain parameters such as detection probability are poorly determined from observed data. Relaxing priors can lead to numerical issues during model estimation for states with fewer cases or noisier data.

Process model

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

The state variables are initialized as:

$$X(0) = \Big(S(0), E(0), I(0), R(0), H(0), D(0), C(0)\Big) = \Big(N - E_0 - I_0 - H_0 - D_0, E_0, I_0, 0, H_0, D_0, I_0\Big)$$

The updates are

$$\beta_{k+1} = \beta_k \times \exp(\epsilon_k), \qquad \epsilon_k \sim \mathcal{N}(0, 0.2)$$
$$X(k+1) = \texttt{ode\_solve}\Big(X(k),\ dX/dt,\ \beta_k\Big)$$

The contact rate β_k undergoes an exponentiated Gaussian random walk starting from β_0 (defined above) with scale τ. The function $\texttt{ode\_solve}$ finds the state vector at time $k+1$ by simulating the ODE for one time step with (constant) contact rate β_k.

Observation model

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

$$y(t) \sim \mathcal{N}(\mu_y, 0.15\mu_y), \qquad \mu_y = pC(t)$$
$$z(t) \sim \mathcal{N}(\mu_y, 0.15\mu_y), \qquad \mu_z = p_d D(t)$$

Our models weight the log-probabilities of the two different types of observations differently to encourage them to focus slightly more on deaths than confirmed cases. Specifically $\log p(y(t)| \cdots)$ is scaled by 0.5 and $\log p(z(t)|\cdot)$ is scaled by 2.0.