v1: 26 May 2020
v2: 1 June 2020

Notes on Parameter Distributions for Monte Carlo simulations of an SIR model, v2

David Kaiser

Center for Theoretical Physics, MIT

We want to sample both the per-capita contact rate β and the average duration of infection γ^{-1} from (semi-)realistic distributions while also keeping the relative width of each distribution comparable, so that the effects on the basic reproductive number $\mathcal{R}_0$ are comparable. We want to keep the average value of $(\mathcal{R}_0)_{\rm avg} = 2.31$ to reflect a global average for analyses of COVID-19. We also want to keep the average duration of infection at $(\gamma^{-1})_{\rm avg} = 7.0$ days. Hence we want to keep the average per-capita rate $\beta_{\rm avg} = (\gamma)_{\rm avg} \times (\mathcal{R}_0)_{\rm avg} = 0.33$.

We will sample the independent parameters β and γ^{-1} from Gamma distributions. The probability density function for the Gamma distribution is given, in general, in the form

$$f_X(x; \mu, \kappa, s) = \frac{1}{\kappa \Gamma(s)} \left(\frac{x - \mu}{\kappa} \right)^{s-1} \exp \left[- \frac{(x - \mu)}{\kappa} \right]. \tag{1}$$

The parameter s is known as the "shape" parameter, and κ is known as the "scale" parameter, while μ sets the minimum value that the parameter X may assume. For the Gamma distribution, we take $\mu, s, \kappa > 0$. According to Wikipedia (!), given the Gamma distribution, the mean value is given by

$$(\bar{x} - \mu) = s\kappa\,, \tag{2}$$

and the variance is given by

$$\text{Var} = s\kappa^2\,. \tag{3}$$

To define the (dimensionless) relative width, I suggest we consider the quantity

$$W_X \equiv \frac{\sqrt{\text{Var}}}{\bar{x}} = \frac{\sqrt{s}\,\kappa}{\mu + s\kappa}. \tag{4}$$

Our challenge is therefore to find values of μ, s, and κ for distributions both for β and γ^{-1} such that the mean values obey $\bar{\beta} = 0.330$ and $\overline{\gamma^{-1}} = 7.0$ days, while also keeping $W_\beta \simeq W_{1/\gamma}$. I suggest the following selections:

$$\begin{aligned} \mu_\beta = 0.21\,,\ \kappa_\beta = 0.01\,,\ s_\beta = 12, \\ \mu_{1/\gamma} = 4.5\,,\ \kappa_{1/\gamma} = 0.25\,,\ s_{1/\gamma} = 10. \end{aligned} \tag{5}$$

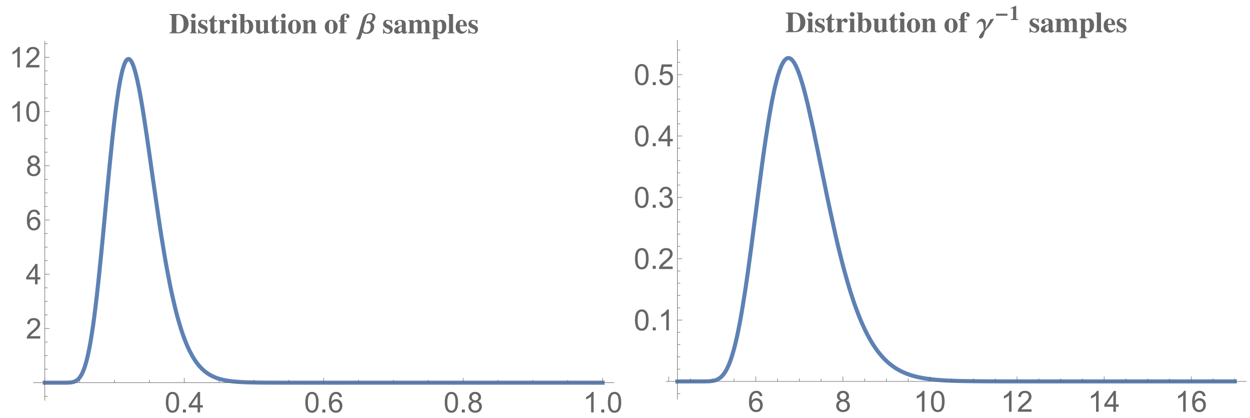

Figure 1: Proposed distributions from which to draw β and γ^{-1} for our Monte Carlo simulations, consistent with the parameter selections in Eq. (5).

With those selections, we find

$$\begin{aligned} \bar{\beta} &= \mu_\beta + s_\beta\, \kappa_\beta = 0.21 + 12 \times 0.01 = 0.330\,, \\ \overline{\gamma^{-1}} &= \mu_{1/\gamma} + s_{1/\gamma}\, \kappa_{1/\gamma} = 4.5 + 10 \times 0.25 = 7.0\,\text{days}\,, \end{aligned} \tag{6}$$

exactly as desired. Moreover, we find

$$\begin{aligned} W_\beta &= \frac{\sqrt{s_\beta}\,\kappa_\beta}{\mu_\beta + s_\beta \kappa_\beta} = 0.105 \simeq 0.11\,, \\ W_{1/\gamma} &= \frac{\sqrt{s_{1/\gamma}}\,\kappa_{1/\gamma}}{\mu_{1/\gamma} + s_{1/\gamma}\,\kappa_{1/\gamma}} = 0.113 \simeq 0.11\,. \end{aligned} \tag{7}$$

The two distributions are shown in Fig. 1.

As a quick reality check, we may calculate the average values of β and γ^{-1} given these selections for the relevant Gamma distributions:

$$\begin{aligned} (\beta)_{\text{avg}} &= \int_{\mu_\beta}^{\infty} dx\, x f_\beta(x; \mu_\beta, \kappa_\beta, s_\beta) = 0.330\,, \\ \left(\gamma^{-1}\right)_{\text{avg}} &= \int_{\mu_{1/\gamma}}^{\infty} dx\, x f_{1/\gamma}(x; \mu_{1/\gamma}, \kappa_{1/\gamma}, s_{1/\gamma}) = 7.00\,\text{days}\,, \end{aligned} \tag{8}$$

exactly as desired. We can also quickly confirm that the variance behaves as expected from Eq. (3). Recall that the variance is defined as

$$\text{Var} = \int_{\mu}^{\infty} dx\, x^2 f_X(x; \mu, \kappa, s) - \left[\int_{\mu}^{\infty} dx\, x f_X(x; \mu, \kappa, s)\right]^2\,. \tag{9}$$

Using the parameters in Eq. (5), I find

$$\begin{aligned} \text{Var}_\beta^{(A)} &= s_\beta\, \kappa_\beta^2 = 1.2 \times 10^{-3}\,, \quad \text{Var}_\beta^{(B)} = 1.2 \times 10^{-3}\,, \\ \text{Var}_{1/\gamma}^{(A)} &= s_{1/\gamma}\, \kappa_{1/\gamma}^2 = 0.625\,, \quad\;\; \text{Var}_{1/\gamma}^{(B)} = 0.625\,, \end{aligned} \tag{10}$$

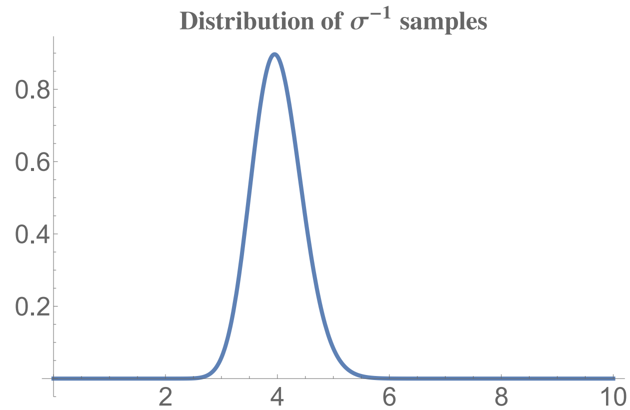

Figure 2: Proposed distribution from which to draw σ^{-1} for our Monte Carlo simulations, consistent with the parameter selections in Eq. (11).

where, for each parameter, $\mathrm{Var}_X^{(A)}$ is calculated according to Eq. (3) and $\mathrm{Var}_X^{(B)}$ is calculated according to Eq. (9).

For the standard SEIR model, we also want to sample the average latency period $1/\sigma$ from within an appropriate distribution. In this case, given our "by-hand" variation of $1/\sigma$ in our early draft, I think it makes sense to draw from a distribution that has a maximum value near 5 days but extends down to shorter periods, while also keeping the relative width $W_{1/\sigma} \simeq 0.11$. I suggest we use:

$$\mu_{1/\sigma} = 0\,, \kappa_{1/\sigma} = 0.05\,, s_{1/\sigma} = 80\,, \tag{11}$$

which yield

$$\overline{1/\sigma} = 4.0\,\mathrm{days}\,, \;\; W_{1/\sigma} = 0.112\,. \tag{12}$$

The probability distribution function for $1/\sigma$ with these parameters is shown in Fig. 2. Using this distribution, I calculate that our random draws for the Monte Carlo simulations should yield $1/\sigma \leq 3.2$ days in about 3.0% of the runs, and should yield $1/\sigma \geq 5$ days in about 1.7% of the runs.

To simulate distinct scenarios with the age-segmented "Cambridge" model, first recall that the full contact matrix is written as in Eq. (28) of our April report:

$$C_{ij}(t) = C_{ij}^H + u^W(t)\,C_{ij}^W + u^S(t)\,C_{ij}^S + u^O(t)\,C_{ij}^O\,, \tag{13}$$

where the superscripts refer to contacts within the home (H), work (W), school (S), and other (O). Policy interventions are encoded in the time-dependent coefficients $u^X(t)$. I propose that for each of the scenarios described below, we draw the parameters u^X from Gamma distributions *once* per run, setting each parameter $u^X(t)$ to the randomly selected *constant* value, beginning and ending at the appropriate time-step of the simulation.

Note from Eq. (4) that for Gamma distributions for which we set the minimum $\mu = 0$, the relative width simplifies:

$$W_X = \frac{\sqrt{s}\,\kappa}{\mu + s\kappa} \longrightarrow \frac{1}{\sqrt{s}} \quad \text{for } \mu = 0\,. \tag{14}$$

To remain consistent with the relative widths used throughout our analysis of the standard SIR and SEIR models, we can fix all random draws to come from distributions such that $W_X = 0.11$.

I propose we try the following. We first infer a best-fit value of $\bar{\beta}$ from the early (no-intervention) data, through day 21 (as we had done in our April report), upon holding $1/\gamma = 7.0$ days fixed and setting all coefficients $u^X = 1$ fixed (again, as in our April report).

For all simulations, we then conduct Monte Carlo runs in which we randomly draw $\bar{\beta}$ from a Gamma distribution with $\mu_{\bar{\beta}} = 0$ and we fix $s_{\bar{\beta}} = 80$ (such that $1/\sqrt{s} = 0.112$), and fix $\kappa_{\bar{\beta}}$ such that the average value of $\bar{\beta}$ drawn from the distribution yields $(\bar{\beta})_{\rm avg} = s_{\bar{\beta}}\,\kappa_{\bar{\beta}}$, where $(\bar{\beta})_{\rm avg}$ is the best-fit value from the early-data calibration. We likewise draw $1/\gamma$ for each run from a Gamma distribution with the same parameters for $1/\gamma$ as in Eq. (5). Then we consider 3 distinct scenarios:

1. *Scenario 1*, no policy interventions: draw $\bar{\beta}$ and $1/\gamma$ from Gamma distributions as described above, while setting each coefficient $u^X(t) \to 1$ for the entire duration of the simulation. Run the simulations for at least 100 days for each realization.

2. *Scenario 2*, a "full lockdown" for 5 weeks (= 35 days), beginning on day 21 and running through day 56. Unlike in the original Cambridge-model paper, we now try to incorporate stochastic noise / compliance effects as follows. For each realization, we again draw $\bar{\beta}$ and $1/\gamma$ from the Gamma distributions described above. In addition, we set all $u^X = 1$ for days 1 - 20 and again set $u^X = 1$ for days 57 - 100. For days 21 - 56, we perform a single random draw per realization for each of the u^X coefficients as follows, treating the coefficients as constants (on each run) between days 21 and 56:

$u^W(t)$ [for days $21 - 56$]: random draw from Gamma distribution with $\mu_W = 0$, $\kappa_W = 0.4/80 = 0.005$, and $s_W = 80$. This yields an average value $(u^W)_{\rm avg} = 0.4$ with relative width $W_W = 1/\sqrt{80} = 0.112$. This is consistent with about 40% of the workforce being categorized as "essential."[1]

[1] According to `https://econofact.org/essential-and-frontline-workers-in-the-covid-19-crisis`, as much as 70% of the US work force has been categorized as "essential"; let's simulate a case in which the average value is large but still below 50%.

$u^S(t)$ [for days $21 - 56$]: random draw from Gamma distribution with $\mu_S = 0$, $\kappa_S = 0.05/80 = 6.25 \times 10^{-5}$, and $s_S = 80$. This yields an average value $(u^S)_{\text{avg}} = 0.05$, with relative width $W_S = 1/\sqrt{80} = 0.112$. This is meant to capture the idea that schools are closed but school-age children might (on occasion) still encounter other children outside the home.

$u^O(t)$ [for days $21 - 56$]: random draw from Gamma distribution with $\mu_O = 0$, $\kappa_0 = 0.15/80 = 0.001875$, and $s_O = 80$. This yields an average value $(u^O)_{\text{avg}} = 0.15$, with relative width $W_O = 1/\sqrt{80} = 0.112$. This is consistent with some people still encountering other people outside the home, some of the time, even post-lockdown.

3. *Scenario 3*, a "partial lockdown" for 5 weeks (= 35 days), beginning on day 21 and running through day 56. In this scenario, we imagine that the government keeps schools closed, but allows (non-essential) businesses to open roughly half time (say, on alternating days of the week). This would correspond to the selections:

$u^W(t)$ [for days $21 - 56$]: random draw from Gamma distribution with $\mu_W = 0$, $\kappa_W = 0.7/80 = 0.00875$, and $s_W = 80$. This indicates that 40% of the workforce is deemed "essential," while the remaining workforce goes back to work half-time.

$u^S(t)$ [for days $21 - 76$]: random draw from Gamma distribution with $\mu_S = 0$, $\kappa_S = 0.05/80 = 6.25 \times 10^{-5}$, and $s_S = 80$, same as for Scenario 1.

$u^O(t)$ [for days $21 - 56$]: random draw from Gamma distribution with $\mu_O = 0$, $\kappa_O = 0.3/80 = 0.00375$, and $s_O = 80$.

As in our analysis of the standard SIR and SEIR models, for all three of these scenarios with the Cambridge model, we can track quantities of interest and compute confidence intervals based on the 68% and 95% quantiles. I suggest we track the number of infected people at three specific times and the total accumulated cases at those same three times for all three scenarios: $\{I(t_{21}), T(t_{21})\}$, $\{I(t_{56}), T(t_{56})\}$, and $\{I(t_{100}), T(t_{100})\}$. Then we can begin to assess whether the likely outcomes from various policy scenarios would lead to likely outcomes that were clearly, statistically significantly distinct from each other.