

## AMSC/MATH 420, Spring 2015 Modeling Epidemics: Team Homework 2a

due Tuesday April 7

In class we discussed the two-group SIR model. Consider now the two-group SI model, setting  $r_1 = r_2 = 0$ . Then since  $N_1 = S_1 + \mathcal{I}_1$  and  $N_2 = S_2 + \mathcal{I}_2$  are constant over time, we can equivalently consider the equations

$$\begin{aligned}d\mathcal{I}_1/dt &= p_{11}(N_1 - \mathcal{I}_1)\mathcal{I}_1 + p_{12}(N_1 - \mathcal{I}_1)\mathcal{I}_2 \\d\mathcal{I}_2/dt &= p_{21}(N_2 - \mathcal{I}_2)\mathcal{I}_1 + p_{22}(N_2 - \mathcal{I}_2)\mathcal{I}_2\end{aligned}$$

Now we have 6 parameters – 4  $p$ 's and 2  $N$ 's – plus the initial conditions  $\mathcal{I}_1(0)$  and  $\mathcal{I}_2(0)$  that determine a particular solution of the equations.

Monthly data on AIDS diagnoses in various metropolitan areas (and the entire U.S.) is available at <http://wonder.cdc.gov/aids-v2002.html>. To download data for a particular geographic area, select “Month Diagnosed” in Section 1 next to “Group Results By”, and select the geographic area in Section 2. You can select a particular demographic group in Section 5, or select a demographic classification (for our purposes, “Sex and Sexual Orientation”) in addition to “Month Diagnosed” for “Group Results By”. Go to Section 8, click “Show Zero Values”, then click one of the “Send” buttons. Once the data appears, you can save the data to a file with the “Export” button, or copy-and-paste. As in the previous project, please use only the first 20 years of data, 1982-2001.

Assign to Group 1 the Male Bisexual and Male Homosexual categories, and to Group 2 the Female and Male Heterosexual categories. (By the way, according to the documentation, pediatric cases are not included in the metropolitan area data.) Note that you can get the same groupings if you download the data by “HIV Exposure Category” instead of demographic classification and assign the Male Homosexual/Bisexual categories to Group 1. Looking at the data by exposure category suggests that at least early in the epidemic, most Group 1 infections were via sexual activity and most Group 2 infections were via IV drug use.

1. Let  $y_1(t)$  and  $y_2(t)$  be the monthly diagnoses for each group in month  $t$  and let  $T$  be the number of months of data. For each of your geographic areas, find parameters minimize the sum of squares of residuals

$$E = \sum_{t=1}^T [(y_1(t) - \mathcal{I}'_1(t))^2 + (y_2(t) - \mathcal{I}'_2(t))^2].$$

Please report what values of the parameters you found and what the RMS error  $\sqrt{E/T}$  is for those parameters, and graph your model solution with these parameters overlaid with the data (separate graphs for Group 1 and Group 2).

2. Simulate the model for different values of the parameters and the initial populations, using values with comparable magnitude to those you get from fitting the data. Look at when the rate of new infections peaks for each group, i.e., when  $d\mathcal{I}_1/dt$  peaks versus when  $d\mathcal{I}_2/dt$  peaks. Can you deduce any general rules about how the parameters affect the time difference between the two peaks?