The spread of an infection through a region

Shortcomings of the *SIR* model

In the original SIR model we have

$$\begin{aligned}\frac{dS}{dt} &= -aIS = -aI(t)\cdot S(t)\\ \frac{dI}{dt} &= aIS - bI = aI(t)\cdot S(t) - bI(t),\end{aligned}$$

where $S(t)$ is the number of susceptibles, and $I(t)$ is the number of infected, at any time t.

We can interpret the first equation as saying that susceptibles fall ill at the per capita rate[1] $aI(t)$ which depends on the number of infected. The multiplier a is a constant that gives us a measure of the effect that infected have on susceptibles. The minus sign indicates that people *leave* the susceptible population as they fall ill.

The second equation has two terms. The first just indicates that losses $-aIS$ from the susceptible population turn into gains $+aIS$ for the infected population. The second term indicates that the infected recover (or are "removed" from the infected population) at the constant per capita rate b.

One of the shortcomings of the original model is that it makes no attempt to indicate that the infected have the greatest influence on the susceptibles *that are geographically closest to them.* In fact, it ignores the possibility that the populations are spread over a region and that it takes time for an infection to spread across that region. We now develop a refinement of the original SIR model that addresses these shortcomings.

Extending the *SIR* model to a region

Let's suppose we want to follow the course of an infection through a city that occupies a region $\mathcal{R}$ in the xy-plane. Let $p = (x, y)$ represent a point (or a vector) in $\mathcal{R}$, and let $\|p\| = \sqrt{x^2 + y^2}$ be the length of this vector. To indicate that S and I now vary from point to point in $\mathcal{R}$, write

$$S = S(t, p) = S(t, x, y) \quad \text{and} \quad I = I(t, p) = I(t, x, y)$$

We interpret $S(t, p)$ to be the *density* of the susceptible population—for example, in persons per square mile—in a small square centered at p at the time t. Interpret $I(t, p)$ and $R(t, p)$ similarly. Using the original expression

---

[1] The per capita rate of change of the population S is S'/S; in this case, $S'/S = -aI$.

$-aIS$ as a model, we will say that the rate at which infected at one point P affect susceptibles at another point p is

$$-a(p,P)I(t,P)\cdot S(t,p).$$

As it must, the multiplier a now depends on the locations of both the susceptible and the infected.

Of course it is the combined effect of the infected $I(t,P)$ at *all* points P that determines the rate at which susceptibles $S(t,p)$ at the single point p fall ill. This combined effect is given by the integral

$$\begin{aligned}\frac{\partial S}{\partial t}(t,p) &= \iint\limits_{\mathcal{R}} -a(p,P)I(t,P)\cdot S(t,p)\,dP \\ &= -\left(\iint\limits_{\mathcal{R}} a(p,P)I(t,P)\,dP\right)\cdot S(t,p).\end{aligned}$$

To get the evolution equation for $I(t,p)$ we'll continue to assume the infected recover at a per capita rate b that is the same at all places (and thus is a constant):

$$\frac{\partial I}{\partial t}(t,p) = \left(\iint\limits_{\mathcal{R}} a(p,P)I(t,P)\,dP\right)\cdot S(t,p) - bI(t,p).$$

Remember that P is a vector (let's say $P=(X,Y)$) so the integral that appears here is a double integral; in terms of the components x, y, X, and Y of p and P, respectively, it has the form

$$\iint\limits_{\mathcal{R}} a(x,y,X,Y)I(t,X,Y)\,dXdY.$$

However, we'll usually keep the vector form because the expression is simpler and usually easier to read.

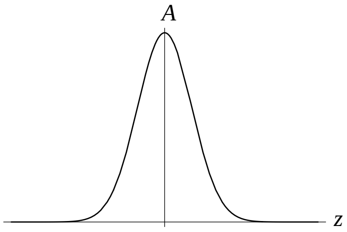
Simplifying the new model

There are several practical steps we can take to simplify these partial differential equations—in particular, we can eventually eliminate the integrals.

Consider first the function $a(p,P)$ that tells us how much an infected at P affects a susceptible at p. It is reasonable to think that the effect shouldn't depend on the particular points P and p, but only on the distance $\|P-p\|$ between them. In other words, we should be able to write

$$a(p,P)=A(\|P-p\|)$$

for some suitable function A of the *single* variable $z=\|P-p\|$. Furthermore, $A(z)$ should be largest when P is near p, that is, when $z=\|P-p\|$ is near 0. Therefore, the graph of $A(z)$ should look something like the graph on the right. In the absence of any exact knowledge about A, let's assume that the effect is zero beyond a certain small distance δ:



$$A(\|P-p\|)=\begin{cases} f(\|P-p\|) & \text{if } \|P-p\|\le\delta \\ 0 & \text{otherwise}\end{cases}$$

Thus

$$\iint\limits_{\mathcal{R}} a(p,P)I(t,P)\,dP=\iint\limits_{\mathcal{R}} A(\|P-p\|)I(t,P)\,dP=\iint\limits_{\mathcal{D}_p} f(\|P-p\|)I(t,P)\,dP.$$

In these integrals, P varies while p is fixed. The domain of the last integral is the disk $\mathcal{D}_p$ of radius δ centered at the point p in the P-plane.[2]

Let's now make the coordinate change $Q=P-p$. This translates the point p in the P-plane to the origin in the Q-plane, so it translates the disk $\mathcal{D}_p$ to the disk $\mathcal{D}_0$ of radius δ centered at the origin. Since $dQ=dP$ and $P=p+Q$, our integral becomes

$$\iint\limits_{\mathcal{D}_0} f(\|Q\|)I(t,p+Q)\,dQ.$$

In this integral, Q is restricted to the disk of radius δ centered at the origin, so $\|Q\|$ is small. We could therefore write $I(t,p+Q)\approx I(t,p)$. Since $I(t,p)$ does not depend upon the integrating variable Q, the integral simplifies to

$$\iint\limits_{\mathcal{D}_0} f(\|Q\|)I(t,p+Q)\,dQ\approx I(t,p)\iint\limits_{\mathcal{D}_0} f(\|Q\|)\,dQ=\alpha I(t,p).$$

---

[2] If p is close to the boundary of $\mathcal{R}$ we must remove from $\mathcal{D}_p$ any part of it that extends outside $\mathcal{R}$.

Here α is the integral of $f(\|Q\|)$ over $\mathcal{D}_0$; it is a positive number whose actual value doesn't concern us. The evolution equation for S becomes

$$\frac{\partial S}{\partial t}(t,p) = -\alpha I(t,p) \cdot S(t,p),$$

and this is essentially the same as the original differential equation—we've gained nothing!

The approximation $I(t, p+Q) \approx I(t,p)$ is, in effect, the zeroth-order Taylor expansion of $I(t,p+Q)$ as a function of Q. We will get a more useful approximation, however, by using the first- or second-order Taylor expansion. If we let $Q=(u,v)$, then the second-order Taylor expansion is

$$\begin{aligned} I(t,x+u,y+v) &\approx I(t,x,y) + I_x(t,x,y)u + I_y(t,x,y)v \\ &\quad + \tfrac{1}{2}(I_{xx}(t,x,y)u^2 + 2I_{xy}(t,x,y)uv + I_{yy}(t,x,y)v^2), \\ \text{or} \quad I(t,p+Q) &\approx I(t,p) + I_x(t,p)u + I_y(t,p)v \\ &\quad + \tfrac{1}{2}(I_{xx}(t,p)u^2 + 2I_{xy}(t,p)uv + I_{yy}(t,p)v^2). \end{aligned}$$

(In a moment we'll see why we don't stop with the first-order approximation.) With this expression for $I(t,p+Q)$ we can approximate the original integral by a sum of six:

$$\iint\limits_{\mathcal{D}_0} f(\|Q\|) I(t,p+Q)\, dQ \approx I(t,p) \iint\limits_{\mathcal{D}_0} f\left(\sqrt{u^2+v^2}\right) dudv +$$

$$I_x(t,p) \iint\limits_{\mathcal{D}_0} uf\left(\sqrt{u^2+v^2}\right) dudv + \cdots + \tfrac{1}{2} I_{yy}(t,p) \iint\limits_{\mathcal{D}_0} v^2 f\left(\sqrt{u^2+v^2}\right) dudv.$$

We put the various factors $I(t,p)$, $\ldots$, $\frac{1}{2}I_{yy}(t,p)$ outside because they do not depend on the integration variables u and v. Since each of the integrals is just a number $a_1, \ldots, a_6$, we can write the expression more simply as

$$a_1 I(t,p) + a_2 I_x(t,p) + a_3 I_y(t,p) + a_4 I_{xx}(t,p) + a_5 I_{xy}(t,p) + a_6 I_{yy}(t,p)$$

Note that a_1 is just α, introduced above. Moreover, we will now show $a_2 = a_3 = a_5 = 0$. (And then, because $a_2 = a_3 = 0$, the first-order terms will disappear; this explains why there is no point stopping with the first-order Taylor approximation.) To calculate the integrals we'll use polar coordinates because the domain of integration has circular symmetry:

$$u = r\cos\theta \qquad v = r\sin\theta \qquad du\,dv = r\,dr\,d\theta \qquad \sqrt{u^2+v^2} = r.$$

In polar coordinates the disk $\sqrt{u^2+v^2} \leq \delta$ is the described by the simple inequalities

$$0 \leq r \leq \delta, \qquad 0 \leq \theta \leq 2\pi.$$

Therefore $a_2 = \displaystyle\iint\limits_{\mathcal{D}_0} uf\left(\sqrt{u^2+v^2}\right)\,dudv = \int_0^{2\pi}\int_0^{\delta} r\cos\theta\, f(r)\, rdrd\theta$

$$= \int_0^{2\pi} \cos\theta\, d\theta \int_0^{\delta} r^2 f(r)\, dr = 0.$$

The value is zero because the integral of $\cos\theta$ is zero. In a_3, a sine function appears instead of a cosine, but the result is still zero. For a_5 we have

$$a_5 = \iint\limits_{\mathcal{D}_0} uvf\left(\sqrt{u^2+v^2}\right)\,dudv = \int_0^{2\pi}\int_0^{\delta} r^2\cos\theta\sin\theta\, f(r)\, rdrd\theta$$

$$= \int_0^{2\pi} \tfrac{1}{2}\sin 2\theta\, d\theta \int_0^{\delta} r^3 f(r)\, dr = 0,$$

because the integral of $\sin 2\theta$ is zero.

It is also true that $a_4 = a_6$. To see this, note first that

$$a_4 = \iint\limits_{\mathcal{D}_0} u^2 f\left(\sqrt{u^2+v^2}\right)\,dudv = \int_0^{2\pi}\int_0^{\delta} r^2\cos^2\theta\, f(r)\, rdrd\theta$$

$$= \int_0^{2\pi} \cos^2\theta\, d\theta \int_0^{\delta} r^3 f(r)\, dr$$

Next, $a_6 = \displaystyle\iint\limits_{\mathcal{D}_0} v^2 f\left(\sqrt{u^2+v^2}\right)\,dudv = \int_0^{2\pi}\int_0^{\delta} r^2\sin^2\theta\, f(r)\, rdrd\theta$

$$= \int_0^{2\pi} \sin^2\theta\, d\theta \int_0^{\delta} r^3 f(r)\, dr.$$

But $\displaystyle\int_0^{2\pi} \cos^2\theta\, d\theta = \int_0^{2\pi} \sin^2\theta\, d\theta = \pi$, so $a_4 = a_6$, as we claimed.

We can now summarize our results in terms of the original integral:

$$\iint\limits_{\mathcal{R}} a(p,P)I(t,P)\,dP \approx \alpha I(t,p) + \beta(I_{xx}(t,p) + I_{yy}(t,p)) = \alpha I + \beta\nabla^2 I$$

where $\alpha = a_1$ and $\beta = a_4 = a_6$ are certain positive constants that depend on the multiplier function $a(p, P) = f(\|P - p\|)$. The actual values of α and β need not concern us. We have also used the Laplacian

$$\nabla^2 I(t,p) = I_{xx}(t,p) + I_{yy}(t,p) = \frac{\partial^2 I}{\partial x^2}(t,p) + \frac{\partial^2 I}{\partial y^2}(t,p)$$

to simplify the expression further. If we now replace the integral in the evolution equations by $\alpha I + \beta \nabla^2 I$, we get

$$\begin{aligned} \frac{\partial S}{\partial t} &= -(\alpha I + \beta \nabla^2 I) S \\ \frac{\partial I}{\partial t} &= (\alpha I + \beta \nabla^2 I) S - bI. \end{aligned}$$

or, more fully,

$$\begin{aligned} \frac{\partial S}{\partial t}(t,p) &= -[\alpha I(t,p) + \beta \nabla^2 I(t,p)] S(t,p) \\ \frac{\partial I}{\partial t}(t,p) &= [\alpha I(t,p) + \beta \nabla^2 I(t,p)] S(t,p) - bI(t,p), \end{aligned}$$

Thus the effect of considering spatial interactions between susceptibles and infected is to add the diffusion term $\beta(\nabla^2 I)S$ to the model.

Difference equations

To solve the evolution equations, we convert them to difference equations in the usual way. If we let $\Delta y = \Delta x$, then the equation

$$S_t = -[\alpha I + \beta(I_{xx} + I_{yy})]S$$

becomes $\dfrac{S(t+\Delta t, x, y) - S(t, x, y)}{\Delta t} =$

$$-\left[\alpha I(t,x,y) + \beta\left(\frac{I(t, x+\Delta x, y) + I(t, x-\Delta x, y)}{(\Delta x)^2} + \right.\right.$$
$$\left.\left. + \frac{I(t, x, y+\Delta y) + I(t, x, y-\Delta y) - 4I(t,x,y)}{(\Delta x)^2}\right)\right] S(t,x,y)$$

If we set $\gamma = \beta/(\Delta x)^2$ and define $I^+(t,x,y) =$

$$I(t, x+\Delta x, y) + I(t, x-\Delta x, y) + I(t, x, y+\Delta y) + I(t, x, y-\Delta y),$$

then we can rewrite the last equation expressing the single "future" value $S(t + \Delta t, x, y)$ in terms of the others:

$$S(t + \Delta t, x, y) = S(t, x, y) - \Delta t\, S(t, x, y)[(\alpha - 4\gamma)I(t, x, y) + \gamma I^{+}(t, x, y)].$$

(We use I^{+} to denote the sum of the values of I at the indicated four points because those four points are arrayed around (t, x, y) in the shape of the plus sign $+$.)

We can make use of what we've already done to write the corresponding difference equation for I:

$$I(t + \Delta t, x, y) = (1 - b\Delta t)I(t, x, y) - (S(t + \Delta t, x, y) - S(t, x, y)).$$

Computer implementation

For simplicity we assume that the region $\mathcal{R}$ is a rectangle

$$a \le x \le b, \qquad c \le y \le d.$$

The assumption that there is no migration in or out of the region translates into the following boundary conditions:

No one crosses the vertical edges $x = a$ or $x = b$: for all t and y,

$$\frac{\partial S}{\partial x}(t, a, y) = \frac{\partial S}{\partial x}(t, b, y) = \frac{\partial I}{\partial x}(t, a, y) = \frac{\partial I}{\partial x}(t, b, y) = 0$$

No one crosses the horizontal edges $y = c$ or $y = d$: for all t and x,

$$\frac{\partial S}{\partial y}(t, x, c) = \frac{\partial S}{\partial y}(t, x, d) = \frac{\partial I}{\partial y}(t, x, c) = \frac{\partial I}{\partial y}(t, x, d) = 0$$