

Odo Diekmann

Mathematisch Instituut
Universiteit Utrecht
Postbus 80010
3508 TA Utrecht
O.Diekmann@math.uu.nl

Vakantiecursus 2002

Aanstekelijkheid gevangen in een getal

Jaarlijks organiseert het Centrum voor Wiskunde en Informatica (CWI) onder auspiciën van de Nederlandse Vereniging van Wiskundeleraars een vakantiecursus voor wiskundeleraars en andere belangstellenden. Bij deze gelegenheid verschijnt steeds een syllabus met teksten bij de voordrachten. Het NAW heeft een serie gestart waarin geselecteerde teksten uit recente syllabi worden geplaatst. Het derde artikel is afkomstig uit de syllabus 'Wiskunde en Gezondheid' bij de Vakantiecursus 2002. Het onderwerp is het modelleren van besmettelijke ziekten. Odo Diekmann is hoogleraar analyse aan de Universiteit van Utrecht. Hij geldt als een van de pioniers op het gebied van de mathematische biologie in Nederland.

Stel een vrouw krijgt, gemiddeld genomen, R_0 dochters. Dan groeit, uitgaande van een constante man-vrouw verhouding, de populatie in n generaties met een factor R_0^n als $R_0 > 1$ (en neemt met deze factor af als $R_0 < 1$). R_0 wordt reproductiegetal genoemd en heeft drempelwaarde 1 in de zin dat het groter of kleiner zijn dan de drempelwaarde beslissend is voor het onderscheid tussen exponentiële groei enerzijds en exponentiële uitsterving anderzijds (ook wanneer we de populatie niet op generatiebasis maar in de reële tijd beschouwen). Het begrip R_0 is zinvol voor mens, dier en plant.

Stel een besmettelijke ziekte, veroorzaakt door een bacterie of virus, doet haar intrede in een populatie van gastheren. Naar analogie van het bovenstaande kunnen we spreken over een 'geboorte' als een gastheer besmet raakt en over 'nakomelingen' als een gastheer de besmetting overdraagt aan andere gastheren. De vergelijking gaat in zoverre mank dat het moeilijker wordt om nakomelingen te krijgen naarmate er al meer 'geboortes' zijn geweest (we gaan er hierbij van uit dat besmetting uiteindelijk resulteert in immuniteit, zoals bij kinderziektes als mazelen en rode hond, en dat de tijdschaal van demografische doorstroom (de vervanging van immune ouden van dagen door vatbare baby's) lang is

ten opzichte van de tijdschaal waarop overdracht van besmetting plaats vindt. Maar, zoals bij een oplaaiend vuur het onvermijdelijke opraken van de brandstof in eerste instantie buiten beschouwing gelaten kan worden, zo kunnen we ook hier de reductie van het vatbare deel van de gastheerpopulatie bewust negeren als we vooral geïnteresseerd zijn in de vraag of ons een epidemische vloedgolf, een uitbraak, te wachten staat of slechts een paar incidentele gevallen. (Wiskundig gezien komt dit neer op het lineairiseren van een in essentie niet-lineair verband.) Daarom is R_0 ook als het over de verspreiding van besmettelijke ziekten gaat een kenmerkende grootte.

Om een besmettelijke ziekte kwijt te raken moet je zorgen dat, in het gedachte-experiment waarbij je als het ware de ziekteverwekker (opnieuw) introduceert in een volledig vatbare populatie, $R_0 < 1$ is. (Het kan dan overigens, zoals in het geval van BSE, nog best lang duren voor de ziekte ook daadwerkelijk van het toneel verdwenen is.) Vandaar dat het effect van tegenmaatregelen (zoals vaccinatie) vaak afgemeten wordt aan de (ingeschatte) invloed op R_0 . Vandaar dat R_0 zo'n belangrijke grootte is. Vandaar dat we graag willen weten hoe R_0 'opgebouwd' is uit verschillende factoren.

Sexueel overdraagbare aandoeningen

Laten we, om de gedachten te bepalen, een seksueel overdraagbare aandoening (SOA) beschouwen, bijvoorbeeld gonorrhoe. Om het ons gemakkelijk te maken beperken we ons tot het strikt homoseksuele mannelijke deel van de populatie. Onder andere verwaarlozen we dus de invloed via biseksuele individuen van en op de heteroseksuele 'buitenwereld'. En nu we toch bezig zijn de waarheid geweld aan te doen, gooien we er een schepje boven op: we nemen aan dat contacten voortdurend wisselen. Met andere woorden: we verwaarlozen het effect van paarvorming.

Zij p de kans op overdracht bij een contact tussen een vatbaar

en een infectieus individu. Zij T de lengte van de infectieuze periode. Zij c het aantal contacten per eenheid van tijd per individu. Dan stellen we

$$R_0 = pcT. \quad (1)$$

De redenering hierachter is dat een besmet individu gedurende de infectieuze periode cT contacten zal hebben die, in de situatie dat er slechts heel weinig besmette individuen rondlopen, allemaal met vatbaren zijn. En van die cT contacten zal een fractie p dus tot besmetting leiden.

Hoewel vele details van de afleiding aanvechtbaar zijn (we komen hier zo op terug) geeft het resultaat inzicht. Als we via voorlichting bereiken dat symptomen sneller onderkend worden (waarna men naar de dokter kan gaan voor behandeling, hetgeen voor gonorrhoe eenvoudig is), zal T dalen. Het feit dat R_0 recht evenredig is met T toont aan dat je daar wat mee opschiet. Condoomgebruik heeft invloed op p en een soortgelijke conclusie volgt. De meest voor de hand liggende conclusie is uiteraard dat R_0 recht evenredig daalt als de graadmeter c voor de mate van seksuele activiteit naar beneden gaat.

In het bovenstaande zijn p , c en T gemiddelde grootheden (die trouwens niet zo gemakkelijk te bepalen zijn uit experimentele en veldgegevens). Bijvoorbeeld, bij sommige individuen verloopt gonorrhoe asymptomatisch en is bijgevolg de infectieuze periode extreem lang. En uiteraard is er een grote mate van variabiliteit in de mate van seksuele activiteit van individuen. Ook R_0 is een gemiddelde. Om het gemiddelde van een produkt gelijk te kunnen stellen aan het produkt van de gemiddelden hebben we onafhankelijkheid nodig. In de praktijk betekent dit: er mag geen aanleiding zijn om afhankelijkheid te veronderstellen. En inderdaad, als we het hebben over asymptomatisch verloop en mate van seksuele activiteit, dan is die aanleiding er niet. Tot op zekere hoogte is dat 'middelen' dus onschuldig, met name als het gaat om individuele verschillen in infectiviteit. Het wordt echter ingewikkeld als het gaat om individuele verschillen die zowel invloed hebben op vatbaarheid als op infectiviteit. En de mate van seksuele activiteit is hiervan een duidelijk voorbeeld: hoe meer contacten je hebt, des te meer kans loop je om de aandoening op te lopen, terwijl je eveneens des te meer nieuwe gevallen zult veroorzaken zodra je zelf infectieus bent. Het is dit 'dubbelop' effect dat in (1) ontbreekt. En het doel van deze bijdrage aan de vakantiecursus is te laten zien hoe met behulp van een extra beetje wiskunde een systematische methode ontwikkeld kan worden om het effect in rekening te brengen.

Heterogeniteit: de kunst van het middelen

Stel de populatie kan, op grond van het 'type' van individuen, opgesplitst worden in twee deelpopulaties ter grootte van, respectievelijk, $N_1 = 10^5$ en $N_2 = 10^7$ individuen. Type 1 heeft $c_1 = 100$ contacten per jaar en type 2 heeft $c_2 = 10$ contacten per jaar. Dan is de verhouding waarin de twee typen voorkomen 1 : 100 en het gemiddeld aantal contacten per jaar is gelijk aan

$$\bar{c} = \frac{c_1 N_1 + c_2 N_2}{N_1 + N_2} = \frac{1}{101} c_1 + \frac{100}{101} c_2 \approx 11.$$

Stel nu dat er geen preferenties zijn voor het type van een part-

ner. Dan zal ieder individu contacten hebben met de twee typen in de verhouding van hun aandeel in het totaal aantal contacten, oftewel in de verhouding $c_1 N_1 : c_2 N_2 = 1 : 10$.

Bijgevolg zullen de twee typen in deze verhouding voorkomen onder de nieuwe gevallen van besmetting. En dus zal het gemiddeld aantal contacten van *besmette* individuen gelijk zijn aan $\bar{c} = (1/11)c_1 + (10/11)c_2 \approx 18$. Met de verkeerde manier van rekenen vinden we dus $R_0 \approx 11pT$ en met de goede manier $R_0 \approx 18pT$, een aanzienlijk verschil.

De definitie van R_0 : de wiskunde van het middelen

In het voorafgaande namen we aan dat het 'type' van een individu niet verandert in de loop van de tijd (hetgeen wel het geval is bij een kenmerk als leeftijd) en dat contacten symmetrisch zijn (zulks in tegenstelling tot de asymmetrie van, bijvoorbeeld, bloedtransfusie). Dat nemen we ook nu weer aan. Maar we laten nu 'preferenties' toe door te stellen dat een fractie q_{11} van de contacten van type 1 individuen met type 1 individuen is en een fractie q_{22} van de contacten van een type 2 individu met type 2 individuen. Er is dan een consistentie voorwaarde:

$$(1 - q_{11})c_1 N_1 = (1 - q_{22})c_2 N_2.$$

Immers, een fractie $1 - q_{11}$ van de totale hoeveelheid $c_1 N_1$ contacten per tijdseenheid van type 1 individuen is met type 2 individuen en een fractie $1 - q_{22}$ van de totale hoeveelheid $c_2 N_2$ contacten per tijdseenheid van type 2 individuen is met type 1 individuen en deze hoeveelheden betreffen dus allebei de contacten per tijdseenheid waarbij van beide typen een individu betrokken is, dus moeten ze aan elkaar gelijk zijn. (In de vorige sectie was $q_{11} = c_1 N_1 / (c_1 N_1 + c_2 N_2)$ en $q_{22} = c_2 N_2 / (c_1 N_1 + c_2 N_2)$ en is dus inderdaad aan deze voorwaarde voldaan.) Onze taak is R_0 te bepalen in afhankelijkheid van N_1 , N_2 , c_1 , c_2 , q_{11} , q_{22} , p en T .

Het is nu handig de intensiteit en het patroon van de contacten te rangschikken in een matrix

$$\begin{pmatrix} q_{11}c_1 & (1 - q_{22})c_2 \\ (1 - q_{11})c_1 & q_{22}c_2 \end{pmatrix}.$$

Op plaats ij staat hier het aantal contacten per tijdseenheid dat een type j individu heeft met type i individuen (een terzijde voor alle duidelijkheid: een contact tussen een type 1 individu en een type 2 individu wordt zowel geteld onder de index 12 als onder de index 21 en ook onder de indices 11 en 22 worden contacten dubbel geteld omdat we onze boekhouding baseren op individuen en niet op paren van individuen). Na vermenigvuldiging met de factor pT is dit de volgende-generatie-matrix

$$\mathbf{M} = \begin{pmatrix} pTq_{11}c_1 & pT(1 - q_{22})c_2 \\ pT(1 - q_{11})c_1 & pTq_{22}c_2 \end{pmatrix}$$

in de zin dat m_{ij} gelijk is aan het gemiddelde aantal type i nakomelingen van een besmet individu van type j . We hebben nu dus vier R_0 -achtige getallen gerangschikt in een matrix en moeten daar door op de goede manier te middelen één getal van zien te maken dat de naam R_0 verdient.

Als intermezzo stellen we de vraag: hoe 'groot' is een matrix?

Een 2×2 matrix geeft een rekenrecept om een punt in het vlak (een vector) af te beelden op een ander punt. Zo beeldt de matrix

$$\mathbf{K} = \begin{pmatrix} 0 & 10 \\ \frac{1}{1000} & 0 \end{pmatrix}$$

de eenheidsvector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ af op de vector $\begin{pmatrix} 10 \\ 0 \end{pmatrix}$ met lengte 10, dus, in zekere zin, is het 'vergroten' effect een factor 10. En inderdaad, de standaard manier om de *norm* van \mathbf{K} te definiëren is $\|\mathbf{K}\| = \max_{\|x\|=1} \|\mathbf{K}x\|$ en voor dit voorbeeld volgt dan $\|\mathbf{K}\| = 10$. Echter, als we \mathbf{K} nogmaals toepassen wordt $\begin{pmatrix} 10 \\ 0 \end{pmatrix}$ afgebeeld op $\begin{pmatrix} 0 \\ \frac{1}{100} \end{pmatrix}$. In twee stappen is de vector dus gekrompen met een factor $1/100$. En omdat

$$\mathbf{K}^2 = \begin{pmatrix} 0 & 10 \\ \frac{1}{1000} & 0 \end{pmatrix} \begin{pmatrix} 0 & 10 \\ \frac{1}{1000} & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{100} & 0 \\ 0 & \frac{1}{100} \end{pmatrix} \quad \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

geldt dat laatste zelfs voor elke keuze van een startvector. Een factor $1/100$ per twee stappen is een factor $1/10$ per stap. Te bewijzen valt dat $\lim_{n \rightarrow \infty} \|\mathbf{K}^n\|^{\frac{1}{n}}$ bestaat en gelijk is aan $1/10$. De door \mathbf{K} beschreven afbeelding heeft dus de eigenschap dat bij alsmat herhaald toepassen de punten gemiddeld per stap een factor $1/10$ dichter bij de oorsprong komen te liggen. Als we geïnteresseerd zijn in het 'vergroten' (of, zoals in dit geval, 'verkleinend') effect bij herhaald toepassen dan is het juiste antwoord $1/10$ en niet 10. Het uitrekenen van $\|\mathbf{K}^n\|^{\frac{1}{n}}$ is hier eenvoudig, maar in het algemeen erg bewerkelijk. De vraag rijst of er een snellere manier is om die factor $1/10$ op het spoor te komen?

Einde intermezzo, terug naar SOA. Als we de coördinaten x_1 en x_2 interpreteren als de aantallen besmette individuen van het met de index overeenkomende type in een bepaalde generatie, dan vinden we deze aantallen in de volgende generatie door de matrix \mathbf{M} los te laten op de vector x . De matrix is dus weer op te vatten als een rekenrecept om, gegeven een punt in het vlak, een nieuw punt te vinden. Het herhalen van deze afbeelding wordt beschreven door het iteratieschema $x(n+1) = \mathbf{M}x(n)$ waarbij het gehele getal n de generaties nummert.

Stel dat de twee typen van generatie op generatie in een vaste verhouding $x_1 : x_2$ zouden voorkomen, dan zouden we weer een gemiddelde \bar{c} kunnen uitrekenen volgens

$$\bar{c} = \frac{x_1 c_1 + x_2 c_2}{x_1 + x_2} \quad (2)$$

en vervolgens $R_0 = pT\bar{c}$ kunnen nemen. De verhouding correspondeert met een lijn door de oorsprong en dus met een richting in het (x_1, x_2) vlak. De door \mathbf{M} beschreven afbeelding beeldt lijnen door de oorsprong op lijnen door de oorsprong af. We zijn op zoek naar de uitzonderlijke lijn die op zichzelf wordt afgebeeld omdat dan, zoals hierboven uitgelegd, de verhoudingen dezelfde blijven, hetgeen het middelen recht toe recht aan maakt. Dit betekent dat we op zoek zijn naar vectoren x waarvoor $\mathbf{M}x$ een veelvoud van x is, oftewel $\mathbf{M}x = \lambda x$ met λ nog nader te bepalen. Dergelijke x heten eigenvectoren van \mathbf{M} behorende bij de zogeheten eigenwaarde λ (merk op dat als x een eigenvector is dat dan hetzelfde geldt voor elk veelvoud van x ; belangrijk is slechts de door x bepaalde richting, niet de grootte van x). Hoewel conceptueel eigenvectoren en eigenwaarden gelijktijdig op het toneel

verschijnen, is het rekentechnisch mogelijk eerst de eigenwaarden te bepalen en pas daarna, desgewenst, de eigenvectoren. Zonder veel moeite bewijst men dat λ een eigenwaarde van de 2×2 matrix \mathbf{M} is dan en slechts dan als λ voldoet aan de kwadratische karakteristieke vergelijking $\lambda^2 - S\lambda + D = 0$.

Hierbij is S het spoor en D de determinant van \mathbf{M} (expliciet gegeven door $S = m_{11} + m_{22}$ en $D = m_{11}m_{22} - m_{12}m_{21}$). Er zijn dus, uitzonderingen daargelaten, precies twee eigenwaarden. Welke van de twee moeten we nu hebben?

Er zijn verschillende manieren om het onderscheid tussen de twee eigenwaarden nader uit te werken en zodoende tot een antwoord te komen. Gelukkig leveren die verschillende manieren hetzelfde antwoord op! De diagonaal matrix

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

heeft als prettige eigenschap dat de invariante lijnen precies overeen komen met de coördinaatassen. Op de x_1 -as worden afstanden met een factor λ_1 vergroot (als $\lambda_1 > 1$) en op de x_2 -as met een factor λ_2 . Als $\lambda_2 > \lambda_1 > 1$ en we beginnen in een willekeurig punt van het (x_1, x_2) vlak en passen vervolgens alsmat deze afbeelding toe, dan geldt $x_1(n) = \lambda_1^n x_1(0)$, $x_2(n) = \lambda_2^n x_2(0)$ zodat

$$\frac{x_1(n)}{x_2(n)} = \left(\frac{\lambda_1}{\lambda_2} \right)^n \frac{x_1(0)}{x_2(0)}$$

De groei van x_1 in de loop der generaties is, relatief ten opzichte van de groei van x_2 , te verwaarlozen en de lijn door de oorsprong waarop het n -de punt zich bevindt kruipt, naarmate n groter wordt, steeds dichter tegen de x_2 -as aan. Asymptotisch voor $n \rightarrow \infty$ geldt, mits $x_2(0) \neq 0$, dat $x_1(n)/x_2(n) \rightarrow 0$ en dat de lengte van $x(n)$ in iedere generatie-stap een factor λ_2 groter wordt.

Het antwoord is dus: we moeten de grootste van de twee eigenwaarden hebben. Wat we precies met de grootste bedoelen is duidelijk als ze allebei positief zijn. In het algemeen kunnen eigenwaarden zeer wel negatief of zelfs complex zijn. De algemene karakterisering is daarom: de grootste in absolute waarde. Uiteraard is het mogelijk dat eigenwaarden dezelfde absolute waarde hebben (dat doet zich bijvoorbeeld voor bij de hierboven geïntroduceerde matrix \mathbf{K}). Bovendien moeten we ons realiseren dat de interpretatie vereist dat de eigenvector x positief is, in de zin dat de componenten x_1 en x_2 niet-negatief zijn. Met andere woorden, de invariante lijn die we zoeken moet door het eerste en derde kwadrant lopen en de bijbehorende eigenwaarde moet positief zijn. Welke garantie is er dat er een welgedefinieerde grootste eigenwaarde is en dat deze deze eigenschappen heeft?

Die garantie is er, en wel in de vorm van een stelling van Perron en Frobenius:

Stelling. Zij \mathbf{M} een $n \times n$ -matrix. Veronderstel dat \mathbf{M} positief is, in de zin dat alle elementen groter of gelijk aan nul zijn. Veronderstel bovendien dat \mathbf{M} primitief is, hetgeen wil zeggen dat voor een zekere macht k alle elementen van de matrix \mathbf{M}^k strikt positief zijn. Dan is er precies één eigenwaarde met een bijbehorende eigenvector die positief is en alle andere eigenwaarden zijn strikt kleiner in absolute waarde.

Men spreekt van de ‘dominante’ eigenwaarde. De bijbehorende eigenvector is zelfs strikt positief onder de genoemde voorwaarden. De hier geformuleerde versie is niet de meest algemene. Een positieve matrix heet *irreducibel* als voor elk paar ij met $1 \leq i, j \leq n$ er een macht k te vinden is (die best van de beschouwde i, j mag afhangen) zo dat het ij -de element van \mathbf{M}^k strikt positief is. Ook voor een irreducibele matrix geldt dat er precies één eigenwaarde is met een positieve eigenvector. Maar, tenzij de matrix in feite primitief is, zijn er andere eigenwaarden met dezelfde absolute waarde. Een 2×2 -matrix als

$$\mathbf{K} = \begin{pmatrix} 0 & + \\ + & 0 \end{pmatrix}$$

treedt op bij een SOA model waarbij de populatie wordt onderverdeeld in mannen en vrouwen en waarbij men alleen strikt heteroseksuele contacten beschouwt. En, evenzo, bij malaria, waar de parasiet van mens naar mug en van mug naar mens overgaat. Dergelijke matrices zijn irreducibel maar niet primitief. De tweede eigenwaarde is negatief, maar heeft dezelfde absolute waarde als de eerste.

Uit de interpretatie volgt dat alle elementen van de volgende-generatie-matrix \mathbf{M} niet-negatief zijn. Het al dan niet primitief zijn van \mathbf{M} heeft te maken met kieskeurigheid. Wanneer voor een willekeurige keuze van een tweetal, niet noodzakelijk verschillende, typen de overeenkomstige deelpopulaties contacten met elkaar onderhouden, dan zijn alle elementen van \mathbf{M} strikt positief en is \mathbf{M} dus zeker primitief.

Deze hele paragraaf was tot nu toe een lange inleiding tot, en motivatie van, de volgende uitspraak: \mathbf{R}_0 is, per definitie, de dominante eigenwaarde van de volgende-generatie-matrix \mathbf{M} .

Tijdens de motivatie gebruikten we formule 2 met x de bij R_0 horende eigenvector, als een stap in de berekening van R_0 . Echter, achteraf bezien is dat veel te omslachtig aangezien we eerst R_0 berekenen uit de karakteristieke vergelijking en pas daarna, eventueel, de eigenvector. De formule is dus meer een hulpmiddel bij het begrijpen van de verbanden, het hoe en waarom, dan een rekenmethode.

Nogmaals het SOA voorbeeld

Gewapend met een algemene definitie van R_0 die tevens een constructief rekenrecept levert, zullen we nu voor een algemene versie van het SOA voorbeeld wat conclusies afleiden.

Stel we hebben niet twee maar n verschillende typen individuen, gekarakteriseerd door het aantal contacten c_j , $j = 1, \dots, n$, per eenheid van tijd. Stel een fractie q_{ij} van deze contacten is met type i individuen, waarbij voldaan is aan de normalisering

$$\sum_{i=1}^n q_{ij} = 1 \tag{3}$$

en de consistentie condities

$$q_{ij}c_jN_j = q_{ji}c_iN_i \tag{4}$$

waarbij N_j de grootte van de deelpopulatie van type j individuen is. Dan heeft de matrix \mathbf{M} elementen $m_{ij} = pTq_{ij}c_j$. In het speciale geval dat er geen enkele voorkeur in het spel is ligt de keuze

$q_{ij} = c_iN_i / \sum_l c_lN_l$ voor de hand. De matrix elementen zijn dan van de vorm $m_{ij} = b_i c_j$ met $b_i = pT / (\sum_l c_lN_l) c_iN_i$ hetgeen betekent dat \mathbf{M} alle lijnen door de oorsprong afbeeldt op de ene lijn waarop b ligt. Immers

$$(\mathbf{M}x)_i = b_i \sum_{j=1}^n c_j x_j \Rightarrow \mathbf{M}x = \left(\sum_{j=1}^n c_j x_j \right) b$$

In het bijzonder volgt uit

$$\mathbf{M}b = \left(\sum_{j=1}^n c_j b_j \right) b \quad \text{dat} \quad R_0 = \sum_{j=1}^n c_j b_j = pT \frac{\sum_{j=1}^n c_j^2 N_j}{\sum_{l=1}^n c_l N_l}$$

(Alle overige eigenwaarden zijn nul in dit geval en de meetkundige interpretatie werkt veel sneller dan de algebraïsche rekenpartij via de karakteristieke vergelijking.) Herschrijven van deze uitdrukking levert extra inzicht op $R_0 = pT (\bar{c} + \bar{c}/\bar{c})$ waarbij

$$\bar{c} = \text{gemiddelde } c = \frac{\sum_{k=1}^n c_k N_k}{\sum_{l=1}^n N_l},$$

$$\bar{c} = \text{variantie van } c = \frac{\sum_{k=1}^n (c_k - \bar{c})^2 N_k}{\sum_{l=1}^n N_l} = \frac{\sum_{k=1}^n c_k^2 N_k - \bar{c}^2}{\sum_{l=1}^n N_l}.$$

Deze uitdrukking laat expliciet zien hoe de fout die we maken door R_0 gelijk te stellen aan $pT \bar{c}$ afhangt van de verdeling van c , met name van de variantie. Merk ook op dat de echte R_0 altijd groter is dan de schatting op basis van een gemiddelde c . Dit brengt ons op het idee dat het verwaarlozen van heterogeniteit wel eens heel algemeen tot een onderschatting van R_0 zou kunnen leiden en nooit tot een overschatting. De volgende stelling toont aan dat dit vermoeden juist is.

Stelling. Zij R_0 de dominante eigenwaarde van de matrix \mathbf{M} met elementen $m_{ij} = pTq_{ij}c_j$ waarbij q_{ij} voldoet aan (3) en (4): dan geldt: $R_0 \geq pT \bar{c}$

Bewijs. Definieer $\mathbf{T} : \mathbf{R}^n \rightarrow \mathbf{R}^n$ door $(\mathbf{T}x)_i = x_i / \sqrt{N_i}$ dan geldt dat de elementen van de matrix \mathbf{TMT}^{-1} gegeven worden door $pTq_{ij}c_j (\sqrt{N_j} / \sqrt{N_i})$ zodat op grond van de consistentie conditie (4), de elementen met de indices ij en ji aan elkaar gelijk zijn. Oftewel, \mathbf{TMT}^{-1} is een symmetrische matrix. Voor een symmetrische matrix \mathbf{S} geldt dat alle eigenwaarden reëel zijn en dat de grootste gelijk is aan $\max_{\|w\|=1} (w, \mathbf{S}w)$ waarbij $(v, w) = \sum_{i=1}^n v_i w_i$ en $\|w\| = \sqrt{(w, w)}$. De vector met de componenten

$$v_i = \frac{\sqrt{N_i}}{\sqrt{\sum_{k=1}^n N_k}}$$

voldoet aan $\|v\| = 1$ en verder geldt, mede op grond van (3) dat

$$\begin{aligned} (v, \mathbf{TMT}^{-1}) &= pT \sum_{i=1}^n \frac{\sqrt{N_i}}{\sqrt{\sum_{k=1}^n N_k}} \sum_{j=1}^n q_{ij}c_j \frac{\sqrt{N_j}}{\sqrt{N_i}} \frac{\sqrt{N_j}}{\sqrt{\sum_{k=1}^n N_k}} \\ &= pT \frac{\sum_{j=1}^n c_j N_j}{\sum_{k=1}^n N_k} = pT \bar{c} \end{aligned}$$

□

Het belang van een wiskundige formulering

Op zich is R_0 een grootte die binnen een populatie-dynamische dan wel epidemiologische context zinvol geïntroduceerd kan worden. Als onderdeel van de definitie komt dan het woordje 'gemiddeld' voor en daarin kan een zekere subtiliteit schuil gaan. Met name is dat het geval wanneer de verdeling waarover gemiddeld moet worden niet a priori vast staat, maar juist mede bepaald wordt door de generatie-op-generatie dynamica die we willen modelleren. Op dat moment levert de wiskundige invalshoek een systematische methode: de onbekende verdeling wordt gekarakteriseerd als de dominante eigenvector van een positieve matrix en R_0 is de bijbehorende eigenwaarde (zodat we desgewenst R_0 kunnen uitrekenen op basis van de matrixelementen zonder ook daadwerkelijk de verdeling uit te rekenen). Door specifieke eigenschappen van de matrix te benutten kunnen we van alles en nog wat aan de weet komen over de manier waarop R_0 bepaald wordt door de ingrediënten van het model.

Daarnaast is er de praktijk. Recentelijk had Nederland te kampen met uitbraken van varkenspest en mond-en-klauwzeer. Op intensive care eenheden van ziekenhuizen zijn resistente bacteriën in opmars. Geslachtsziekten, die sterk teruggedrongen werden toen het besef van de verschrikkingen van AIDS het condoomgebruik omhoog stuwde, nemen weer toe, hetgeen te denken en te vrezen geeft. Wereldwijd zijn tuberculose en malaria in snel tempo het verloren terrein aan het heroveren. Wiskundige modellering is één van de vele onderdelen van de strijd tegen besmettelijke ziekten. Het aspect van 'conceptuele verheldering' stond centraal in dit artikel. Ook in brainstormsessies met veterinaire deskundigen en medici is dikwijls de belangrijkste bijdrage gelegen in het feit dat een wiskundige formulering dwingt tot helderheid over wat voorondersteld wordt met betrekking tot de

aard van de variabelen en hun onderlinge verbanden.

De volgende voorbeelden, van werk waar de auteur van dit hoofdstuk in de loop van vele jaren bij betrokken was, illustreren dit algemene patroon.

- De snelheid waarmee een landbouwpest zich verspreidt over verschillende ruimteschalen (een veld, een gebied, een land, een continent) hangt af van de productie, in de loop der tijd, van kiemen en de verplaatsing van deze kiemen door het luchtruim; met wiskundige hulpmiddelen kan men het verband in kaart brengen.
- Zeehonden houden als ze uit het water komen om te rusten min of meer een constante afstand van elkaar; daardoor verandert de populatiedichtheid niet als een epidemie van Phocine Distemper Virus slachtoffers maakt (zoals in 1988 en nu weer); via een wiskundig model valt na te gaan wat dat voor gevolgen heeft voor het percentage dat besmet wordt.
- Naar aanleiding van de varkenspest epidemie is software in ontwikkeling om, op basis van beschikbaar komende gegevens, in te schatten hoeveel bedrijven al wel besmet maar nog niet gedetecteerd zijn.
- In toenemende mate komen data beschikbaar over ziekenhuis bacteriën als MRSA en VRE die resistent zijn voor allerlei antibiotica; de kunst is via data-analyse inzicht te verkrijgen en conclusies te trekken; daarbij zijn wiskundige modellen onontbeerlijk.

Kwalitatief inzicht in ketens van oorzaak en gevolg is de basis voor zowel de 'educated guess' van de expert als het betekenis toekennen aan kwantitatief rekenwerk in het kader van wiskundige modellen. Het feit dat kwantitatief accurate voorspellingen een wensdroom blijven, doet hier niets aan af. ←

Referenties

- 1 H. Andersson, T. Britton, *Stochastic Epidemic Models and Their Statistical Analysis*, Springer Lecture Notes in Statistics 151 Springer, New York, 2000.
- 2 R.M. Anderson, R.M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, 1991.
- 3 N.J.T. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications* Griffin, London, 1975.
- 4 N.G. Becker, *Analysis of Infectious Disease Data* Chapman & Hall, London, 1989
- 5 A.D. Cliff, P. Haggett, *Atlas of Disease Distributions: Analytical Approaches to Epidemiological Data*, Blackwell, London, 1988.
- 6 D.J. Daley, J. Gani, *Epidemic Modelling: an Introduction*, Cambridge University Press, Cambridge, 1999.
- 7 O. Diekmann, J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, Wiley, Chichester, 2000.
- 8 D. Mollison (ed.), *Epidemic Models: Their Structure and Relation to Data*, Cambridge University Press, Cambridge, 1995
- 9 N. Shigesada, K. Kawasaki, *Biological Invasions: Theory and Practice*, Oxford University Press, Oxford, 1997.