

Nicolas Bacaër

met de hulp van

Leo van Wissen, Rik Verdonck, Minus van Baalen,
Patsy Haccou, Frans Jacobs, Femke van den Berg,
Marjolein Lof, Bart Haegeman, Pieter Trapman,
Isabelle Verjans, Babette de Wolff

Een korte geschiedenis van wiskundige populatiodynamica



Een korte geschiedenis van wiskundige populatiodynamica

Nicolas Bacaër

met de hulp van

Leo van Wissen, Rik Verdonck, Minus van Baalen,
Patsy Haccou, Frans Jacobs, Femke van den Berg,
Marjolein Lof, Bart Haegeman, Pieter Trapman,
Isabelle Verjans, Babette de Wolff

Nicolas Bacaër

(*Institut de recherche pour le développement*, nicolas.bacaer@ird.fr)

Leo van Wissen (Nederlands Interdisciplinair Demografisch Instituut KNAW-NIDI, en Rijksuniversiteit Groningen, wissen@nidi.nl)

Rik Verdonck (Universiteit Hasselt, rik.verdonck@uhasselt.be)

Minus van Baalen

(*Centre national de la recherche scientifique*, minus.van.baalen@ens.fr)

Patsy Haccou (p.haccou@gmail.com)

Frans Jacobs (Universiteit Leiden, f.j.a.jacobs@biology.leidenuniv.nl)

Femke van den Berg (*Fera Science Ltd*, femke.vandenberg@fera.co.uk)

Marjolein Lof (Wageningen Universiteit, marjolein.lof@wur.nl)

Bart Haegeman

(*Centre national de la recherche scientifique*, bart.haegeman@sete.cnrs.fr)

Pieter Trapman (Universiteit van Stockholm, ptrapman@math.su.se)

Isabelle Verjans (isabelleverjans1@gmail.com)

Babette de Wolff (Vrije Universiteit Amsterdam, b.wolff@vu.nl)

Omslagfoto's: het Nederlands Collegium en de Stichting Biermans-Lapôte
in de *Cité internationale universitaire de Paris*

Titre original :

Histoires de mathématiques et de populations

© Cassini, Paris, 2008

Pour l'édition néerlandaise :

© Nicolas Bacaër, Paris, 2022

ISBN : 979-10-343-6356-8

Dépôt légal : janvier 2022

Inleiding

Populatiedynamica is het wetenschapsdomein dat op eenvoudige mechanistische wijze de variaties in de tijd van de omvang en samenstelling van biologische populaties, zoals die van mensen, dieren, planten of microorganismen, tracht te verklaren. Het is verwant met, maar toch duidelijk verschillend van, het meer beschrijvende gebied van de populatiestatistiek. Een gemeenschappelijk punt is dat zij uitgebreid gebruik maken van wiskundige taal.

Populatiedynamica bevindt zich op het snijpunt van verschillende gebieden: wiskunde, sociale wetenschappen (demografie), biologie (populatiegenetica en ecologie) en geneeskunde (epidemiologie). Bijgevolg wordt het niet vaak als één geheel voorgesteld, ondanks de gelijkenissen tussen de problemen die in de verschillende toepassingen aan bod komen. Een opmerkelijke uitzondering is het Franse boek *Wiskundige populatietheorieën* van Alain Hillion¹. Maar dit boek presenteert het onderwerp vanuit het oogpunt van de wiskundige, waarbij verschillende soorten modellen worden onderscheiden: discrete-tijdmodellen ($t = 0, 1, 2, \dots$) en continue-tijdmodellen (t is een reëel getal), deterministische modellen (toekomstige toestanden zijn precies bekend als de huidige toestand precies bekend is) en stochastische modellen (waarbij waarschijnlijkheden een rol spelen). Het boek behandelt vervolgens logisch discrete deterministische modellen, continue deterministische modellen, discrete stochastische modellen en continue stochastische modellen.

In dit boek heb ik getracht hetzelfde onderwerp te bespreken, maar dan vanuit een historisch oogpunt. Onderzoek wordt in zijn context uitgelegd. Korte biografieën van wetenschappers zijn opgenomen. Dit moet het boek leesbaarder maken voor degenen die minder vertrouwd zijn met wiskunde en kan meestal helpen bij het begrijpen van de oorsprong van de bestudeerde problemen. Maar dit boek gaat niet alleen over geschiedenis. Het kan ook dienen als een inleiding tot wiskundige modellering. Het leek belangrijk om de details van de meeste berekeningen op te nemen, zodat de lezer echt de beperkingen van de modellen kan zien. Technische delen worden benadrukt in grijze kaders en kunnen bij eerste lezing worden overgeslagen. In het laatste hoofdstuk wordt aandacht besteed aan de talrijke hedendaagse problemen in de populatiedynamica die men vanuit wiskundig oogpunt kan proberen te

¹ *Presses Universitaires de France*, Parijs, 1986.

analyseren. Voor wie meer wil weten, bevatten de lijsten met referenties aan het eind van elk hoofdstuk ook websites waar originele artikelen kunnen worden gedownload.

Het was niet mogelijk om in een boek van deze lengte een volledig beeld te geven van al het tot nu toe ontwikkelde werk of om alle wetenschappers te bespreken die aan het onderwerp hebben bijgedragen. De gemaakte keuze bevat noodzakelijkerwijs een arbitraire component, met name voor de meest recente decennia. Ik hoop niettemin dat het gekozen staal representatief genoeg is, en dat mensen die op het gebied actief zijn en wier werk niet wordt vermeld, niet teleurgesteld zullen zijn.

Het ideale publiek voor dit boek zou omvatten:

- Studenten van middelbare scholen en universiteiten die zich afvragen welke verbanden er kunnen bestaan tussen de wiskundecursussen die zij moeten volgen en de wereld om hen heen, of studenten die een persoonlijk werkstuk voorbereiden over een thema dat verband houdt met de populatiedynamica.
- Wiskundeleraren die proberen hun cursus aantrekkelijker te maken. De kennis van de vier elementaire bewerkingen is voldoende om de meeste hoofdstukken 1, 2 en 5 te begrijpen. Hoofdstuk 3 kan dienen als inleiding op de toepassingen van logaritmen. Dit boek behandelt ook:
 - recurrente betrekkingen in hoofdstukken 1, 3, 8, 11, 14, 21, 23, 24;
 - differentiaalvergelijkingen in hoofdstukken 4, 6, 12, 13, 16;
 - partiële differentiaalvergelijkingen in hoofdstukken 20, 25;
 - een integraalvergelijking in hoofdstuk 10;
 - toepassingen van de waarschijnlijkheidstheorie in hoofdstukken 2, 7, 8, 9, 15, 16, 17, 18, 19, 22.
- Mensen die reeds vertrouwd zijn met demografie, epidemiologie, genetica of ecologie en hun favoriete discipline willen vergelijken met andere domeinen waarin soortgelijke wiskundige modellen gebruikt worden.
- Lezers die geïnteresseerd zijn in de geschiedenis van de wetenschap.

Voor de Nederlandse vertaling dank ik Leo van Wissen, Rik Verdonck, Minus van Baalen, Patsy Haccou, Frans Jacobs, Femke van den Berg, Marjolein Lof, Bart Haegeman, Pieter Trapman, Isabelle Verjans en Babette de Wolff, die de machinevertaling van *DeepL* hebben gecorrigeerd.

Hoofdstuk 1

De rij van Fibonacci (1202)

In 1202 publiceerde Leonardo van Pisa, later ook bekend onder de naam Fibonacci, een boek dat het Indiase decimale getallenstelsel in Europa populariseerde, dat ook door Arabische wiskundigen was overgenomen. Een van de vele voorbeelden die in het boek worden gegeven heeft betrekking op de groei van een konijnenpopulatie. Het is een van de oudste voorbeelden van een wiskundig model voor de dynamica van een populatie.

Leonardo van Pisa, na zijn dood bekend geworden onder de naam Fibonacci, werd geboren rond 1170 in de Republiek van Pisa toen deze op het hoogtepunt van haar commerciële en militaire macht in de Middellandse Zee wereld was. Rond 1192 werd Fibonacci's vader door de Republiek naar de haven van Bejaia, nu in Algerije, gestuurd om een handelspost te leiden. Zijn zoon voegde zich kort daarna bij hem om zich toe te leggen op het beroep van koopman. Leonardo begon het decimale getallenstelsel te leren dat de Arabieren uit India hadden meegebracht en dat vandaag nog steeds in bijna dezelfde vorm wordt gebruikt: 0, 1, 2, 3, 4, 5, 6, 7, 8 en 9. Terwijl hij voor zaken rond de Middellandse Zee reisde, vergeleek hij de verschillende getallenstelsels en bestudeerde hij de Arabische wiskunde. Terug in Pisa voltooide hij in 1202 een boek in het Latijn, getiteld *Liber abaci* („Boek der berekeningen”), waarin hij het nieuwe getallenstelsel uitlegde en liet zien hoe het gebruikt kon worden voor boekhouding, gewichts- en valutaomrekeningen, rentetarieven en vele andere toepassingen. Hij verzamelde ook de meeste resultaten van de algebra en de rekenkunde die de Arabieren kenden.

Fibonacci behandelde in zijn boek wat men tegenwoordig een probleem uit de populatiedynamica zou noemen. Maar het verscheen gewoon als een rekenoefening te midden van andere ongerelateerde onderwerpen: de vorige sectie in het boek gaat over perfecte getallen, d.w.z. getallen die de som zijn van hun factoren (zoals $28 = 14 + 7 + 4 + 2 + 1$), en de volgende sectie is een probleem over het verdelen van geld onder vier mensen dat equivalent is aan een lineair stelsel van vier vergelijkingen. Hier is een vertaling uit het Latijn van het probleem uit de populatiedynamica:

„Een zeker man hield een konijnenpaar op een bepaalde afgeslo-

ten plaats. Men wenst te weten hoeveel er van deze paren in één jaar worden geschapen, als het in de aard van een paar is om zich na één maand voort te planten en na twee maanden een nieuw konijnenpaar te baren.”

Indien er aan het begin van de eerste maand één paartje pasgeboren konijnen is, zal dit paartje zich na één maand voortplanten, en zal er aan het begin van de tweede maand nog slechts één paartje konijnen zijn. Dit konijnenpaar zal aan het begin van de derde maand een ander konijnenpaar ter wereld gebracht hebben, zodat er dan in totaal twee paren zullen zijn. Het eerste konijnenpaar zal aan het begin van de vierde maand opnieuw een konijnenpaar gebaard hebben. Maar het tweede konijnenpaar zal dan nog geen nieuw konijnenpaar gebaard hebben. Er zullen slechts drie konijnenparen zijn.

Laat P_n het aantal paren konijnen aan het begin van maand n zijn. Het aantal paren konijnen P_{n+1} aan het begin van maand $n + 1$ is de som van het aantal paren P_n aan het begin van maand n en het aantal pasgeboren paren in het begin van maand $n + 1$. Maar alleen de paren konijnen die minstens twee maanden oud zijn baren nieuwe paren konijnen in maand $n + 1$. Dit zijn de paren die er in maand $n - 1$ al waren, en hun aantal is P_{n-1} . Dus

$$P_{n+1} = P_n + P_{n-1}.$$

Dit is een recurrente relatie: deze geeft de grootte (in aantal paren) van de konijnenpopulatie in maand $n + 1$ als functie van de grootten in de twee voorgaande maanden. Vandaar dat Fibonacci gemakkelijk de volgende tabel kon opstellen, waarbij $P_1 = 1$ en $P_2 = 1$.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
P_n	1	1	2	3	5	8	13	21	34	55	89	144	233

In feite beschouwde Fibonacci als beginvoorwaarde de situatie in maand $n = 2$. Aangezien $P_{14} = 144 + 233 = 377$, verkreeg hij uiteindelijk 377 konijnenparen twaalf maanden na zijn beginpunt. Hij merkte op dat deze getalreeks oneindig kon doorgaan.

Na 1202 schreef Fibonacci verschillende andere boeken, zoals *Practica geometriae* in 1220 en *Liber quadratorum* in 1225. Zijn reputatie leidde tot een ontmoeting met keizer Frederik II, die waardering had voor de wetenschap. In 1240 kende de Republiek van Pisa Fibonacci een jaarlijks pensioen toe. Het jaar van zijn dood is onbekend.

Gedurende de volgende eeuwen werd het konijnenprobleem van Fibonacci vergeten en had het geen invloed op de ontwikkeling van wiskundige

modellen voor de dynamica van populaties. Verschillende wetenschappers kwamen in hun studies dezelfde getallenreeks tegen, maar verwezen niet naar Fibonacci of naar een populatie. In verschillende boeken van Kepler staat de opmerking dat de verhouding P_{n+1}/P_n , als n naar oneindig gaat, convergeert naar het gulden getal $\phi = (1 + \sqrt{5})/2$. Dit is een bijzonder geval van een eigenschap die de meeste populatiemodellen gemeen hebben: de neiging om meetkundig toe te nemen (zie de hoofdstukken 3 en 21). In 1728 heeft Daniel Bernoulli, bij het bestuderen van algemene recurrente relaties, de exacte formule

$$P_n = \frac{1}{\sqrt{5}} \left[\frac{1 + \sqrt{5}}{2} \right]^n - \frac{1}{\sqrt{5}} \left[\frac{1 - \sqrt{5}}{2} \right]^n$$

afgeleid. De volledige werken van Fibonacci werden in de negentiende eeuw gepubliceerd. Vanaf dat moment was de reeks (P_n) in boeken over recreatieve wiskunde te vinden onder de naam reeks van Fibonacci.

Het is duidelijk dat, om bovengenoemde populatie konijnen te modelleren, de hypothesen die tot de Fibonacci-reeks leiden verre van realistisch zijn: geen sterfte, geen scheiding van geslachten enz. De belangstelling vanuit de biologie voor deze reeks is in de laatste decennia voortgekomen uit het feit dat verschillende planten structuren bevatten waarin enkele van de getallen P_n voorkomen, bijvoorbeeld 8 en 13 in dennenappels of 34 en 55 in zonnebloemen. Het wetenschappelijk tijdschrift *The Fibonacci Quarterly* is volledig gewijd aan de eigenschappen en toepassingen van de Fibonacci-reeks!

Verdere literatuur

1. Bernoulli, D.: *Observationes de seriebus... Comment. Acad. Sci. Imp. Petropolitanae* 3, 85–100 (1728/1732) → *Die Werke von Daniel Bernoulli*, Band 2, Birkhäuser, Basel, 1982, 49–64.
2. Sigler, L.E.: *Fibonacci's Liber Abaci*. Springer (2002).
3. Vogel, K.: Leonardo Fibonacci. In: Gillespie, C.C. (ed.) *Dictionary of Scientific Biography*, vol. 4, 604–613. Scribner, New York (1971)

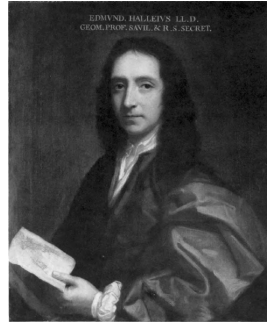
Hoofdstuk 2

Halley's sterftetabel (1693)

In 1693 bestudeerde de beroemde Engelse astronoom Edmond Halley de geboorte- en sterftcijfers van de stad Breslau, die door Caspar Neumann aan de *Royal Society* waren doorgegeven. Hij stelde een sterftetabel op die aangaf hoeveel mensen van een cohort dat in hetzelfde jaar geboren was, een bepaalde leeftijd hadden bereikt. Hij gebruikte zijn tabel ook om de prijs van lijfrentes te berekenen. Dit hoofdstuk brengt dit werk in herinnering en plaatst het in de context van Halleys leven en van de vroege ontwikkelingen van de „politieke rekenkunde” en de waarschijnlijkheidsleer, waarin mensen als Graunt, Petty, De Witt, Hudde, Huygens, Leibniz en de Moivre geïnteresseerd waren.

Edmond Halley werd in 1656 in de buurt van Londen geboren. Zijn vader was een rijke zeepmaker. Edmond raakte al op jonge leeftijd geïnteresseerd in astronomie. Hij ging studeren aan het *Queen's College* van de universiteit van Oxford. Toen in 1675 het Greenwich Observatorium werd ingewijd, kon Halley al terecht bij Flamsteed, de eerste *Astronomer Royal*. Hij onderbrak zijn studie van 1676 tot 1678 om naar het eiland Sint-Helena te gaan en een catalogus op te stellen van de sterren die te zien zijn vanaf het zuidelijk halfrond. Bij zijn terugkeer in Engeland werd hij lid van de *Royal Society*. Hij publiceerde ook de waarnemingen die hij tijdens zijn reis naar Sint-Helena had gedaan over de circulatie van de winden. In 1684 bezocht hij Newton in Cambridge om het verband te bespreken tussen Keplers wetten van de beweging van de planeten en de aantrekkingskracht van de zon. Hij moedigde Newton aan om het later beroemd geworden boek *De wiskundige beginselen van de natuurfilosofie* te schrijven, een boek dat Halley uiteindelijk op eigen kosten publiceerde. Hij was toen werkzaam als klerk van de *Royal Society*. In 1689 ontwierp hij een duikerklok, die hij zelf testte.

Ongeveer in dezelfde tijd verzamelde Caspar Neumann, een theoloog die in Breslau woonde, gegevens over het aantal geboorten en sterfgevallen in zijn stad. Breslau (het huidige Wrocław in Polen) behoorde destijds tot het Habsburgse keizerrijk. De gegevens bevatten ook de leeftijd waarop mensen waren overleden. De informatie kon dus worden gebruikt om een sterftetabel op te stellen, met daarin vermeld de kans op overleven tot een bepaalde leeftijd.



Figuur 2.1:
Halley (1656–1742)

De eerste sterftetabel werd in 1662 in Londen gepubliceerd in een boek getiteld *Natuurlijke en politieke observaties gebaseerd op de lijsten van sterfte*. Dit boek wordt gewoonlijk beschouwd als de grondtekst van zowel de statistiek als de demografie en heeft een vreemde bijzonderheid: men vraagt zich tegenwoordig nog steeds af of het geschreven is door John Graunt, een Londense koopman en auteur die op de omslag van het boek wordt genoemd, of door zijn vriend William Petty, één van de oprichters van de *Royal Society*.

In ieder geval probeerde de sterftetabel in het boek gebruik te maken van de bulletins die sinds het begin van de zeventiende eeuw regelmatig de begravenissen en dopen in Londen meldden. Deze bulletins werden vooral gebruikt om de mensen te informeren over de terugkerende pestepidemieën. Daarom gaven zij de doodsoorzaak aan en niet de leeftijd waarop men stierf. Om een sterftetabel te verkrijgen die de overlevingskansen als functie van de leeftijd aangeeft, moesten Graunt of Petty gissen hoe de verschillende doodsoorzaken zich verhielden tot leeftijdsgroepen. Hun sterftetabel kon hierdoor mogelijk onderhevig zijn aan grote fouten. Het boek was niettemin zeer succesvol, met vijf edities tussen 1662 en 1676. Verschillende steden in Europa waren begonnen met het publiceren van soortgelijke bulletins als die van Londen.

Het was dus bijna dertig jaar na deze eerste sterftetabel dat Neumann, op aanraden van Leibniz, aan Henry Justel, de secretaris van de *Royal Society*, zijn demografische gegevens van de stad Breslau voor de jaren 1687–1691 toezond. Justel stierf kort daarna en Halley kreeg de gegevens in handen, analyseerde ze en publiceerde in 1693 zijn conclusies in de *Philosophical Transactions of the Royal Society*. Zijn artikel is getiteld *Een schatting van de mate van sterfte van de mensheid, verkregen uit merkwaardige tabellen van de geboorten en begravenissen in de stad Breslaw, met een poging om de prijs van lijfrentes op levens te bepalen*.

Voor de bestudeerde periode van vijf jaar stelde Halley vast dat het aan-

tal geboorten in Breslau min of meer gelijk was aan het aantal sterfgevallen, zodat de totale bevolking vrijwel constant was. Om de analyse te vereenvoudigen ging hij ervan uit dat de bevolking zich precies in een stabiele toestand bevond: het jaarlijkse aantal geboorten (aangeduid met P_0), de totale bevolking, het aantal mensen met een leeftijd van k jaar (P_k) en het jaarlijkse aantal sterfgevallen van mensen met een leeftijd van k jaar (D_k) zijn allen constant in de loop van de tijd. Dit benadrukt een bijkomende interessante eigenschap van de gegevens uit Breslau, want een dergelijke vereenvoudiging zou niet mogelijk zijn geweest voor een snel groeiende stad als Londen, waar de statistieken ook nog vertekend werden door de bevolkingsinstroom vanuit het platteland.

Tabel 2.1: De sterftetabel van Halley met de bevolking P_k van k jaar.

k	P_k	k	P_k	k	P_k	k	P_k	k	P_k	k	P_k
1	1.000	15	628	29	539	43	417	57	272	71	131
2	855	16	622	30	531	44	407	58	262	72	120
3	798	17	616	31	523	45	397	59	252	73	109
4	760	18	610	32	515	46	387	60	242	74	98
5	732	19	604	33	507	47	377	61	232	75	88
6	710	20	598	34	499	48	367	62	222	76	78
7	692	21	592	35	490	49	357	63	212	77	68
8	680	22	586	36	481	50	346	64	202	78	58
9	670	23	579	37	472	51	335	65	192	79	49
10	661	24	573	38	463	52	324	66	182	80	41
11	653	25	567	39	454	53	313	67	172	81	34
12	646	26	560	40	445	54	302	68	162	82	28
13	640	27	553	41	436	55	292	69	152	83	23
14	634	28	546	42	427	56	282	70	142	84	20

De gegevens van Breslau lieten een gemiddelde van 1.238 geboorten per jaar zien: dit is de waarde die Halley nam voor P_0 . In principe kon hij uit de gegevens ook het jaargemiddelde D_k berekenen van het aantal sterfgevallen onder mensen met leeftijd k , voor alle $k \geq 0$. Met de formule

$$P_{k+1} = P_k - D_k, \quad (2.1)$$

kon hij een tabel samenstellen voor de waarden van P_k . Omgekeerd kan men de waarden van D_k die hij gebruikte vinden uit de formule $D_k = P_k - P_{k+1}$: $D_0 = 238$, $D_1 = 145$, $D_2 = 57$, $D_3 = 38$ enzovoort. In feite heeft Halley zijn resultaten een beetje herschikt, hetzij om ronde getallen te krijgen (dit is het geval voor de waarde van D_1 , die lichtjes gewijzigd is zodat $P_1 = 1.000$),

hetzij om bepaalde onregelmatigheden glad te strijken die te wijten zijn aan de kleine aantallen sterfgevallen op hoge leeftijd in een vijfjarige studie. Door de som van alle getallen P_k in de tabel te nemen, verkrijgt Halley¹ een schatting van de totale bevolking van Breslau van bij benadering 34.000. Samengevat had deze methode het grote voordeel dat er geen algemene volkstelling nodig was, maar alleen kennis van het aantal geboorten en sterfgevallen en van de leeftijden waarop mensen sterven gedurende een periode van enkele jaren.

De sterftetabel van Halley diende als referentie voor verschillende werken in de achttiende eeuw (zie hoofdstuk 4). Hoewel de waarden van P_k specifiek waren voor de stad Breslau, zou men kunnen stellen dat de verhouding P_{k+1}/P_k de kans is om tot de leeftijd van $k + 1$ te overleven in de wetenschap dat men de leeftijd van k reeds heeft bereikt. Men zou bijvoorbeeld kunnen verwachten dat een kind van één jaar een kans van 661 op 1.000 heeft om de leeftijd van 10 jaar te bereiken, of een kans van 598 op 1.000 om de leeftijd van 20 jaar te bereiken.

Halley gebruikte zijn sterftetabel ook om de prijs van lijfrentes op levens te berekenen. In de zestiende en zeventiende eeuw hadden verschillende steden en staten dergelijke lijfrentes aan hun burgers verkocht om aan geld te komen. De kopers ontvingen elk jaar tot aan hun dood een vast geldbedrag, dat gelijk was aan een bepaald percentage van de oorspronkelijk betaalde som, vaak het dubbele van de toenmalige rentevoet, maar onafhankelijk van de leeftijd van de koper. Uiteraard liep de instelling het risico failliet te gaan als te veel mensen met een zeer lange levensverwachting deze lijfrentes kochten. Dit probleem kon niet correct worden aangepakt zonder een betrouwbare sterftetabel.

In 1671 hadden Johan De Witt, raadpensionaris van Holland, en Johannes Hudde, later burgemeester van de stad Amsterdam, al nagedacht over het probleem van de berekening van de prijs van lijfrentes. Zij vreesden een invasie van Franse troepen en wilden geld inzamelen om het leger te versterken. Zij beschikten over gegevens over mensen die tientallen jaren eerder lijfrentes op levens hadden gekocht, met name de leeftijd waarop de lijfrentes waren gekocht en de leeftijd waarop de mensen waren overleden. Zij slaagden erin de prijs van lijfrentes min of meer correct te berekenen, maar hun methode werd later vergeten. Het jaar daarop werd Holland binnengevallen en werd De Witt door een menigte gelyncht.

Halley bekeek het probleem in 1693 opnieuw met de sterftetabel van Breslau en uitgaande van een rentevoet van 6%. De berekeningsmethode is eenvoudig. Stel i is de rentevoet. Laat R_k de prijs zijn waartegen een per-

¹Voor mensen ouder dan 84 jaar heeft Halley eenvoudigweg verondersteld dat hun aantal 107 is.

soon van leeftijd k een lijfrente van, zeg, een pond per jaar kan kopen. Deze persoon heeft een kans van P_{k+n}/P_k om op de leeftijd van $k+n$ nog in leven te zijn. Het pond dat de staat belooft uit te keren als hij die leeftijd bereikt, kan worden verkregen door $1/(1+i)^n$ pond van de beginsom te storten tegen de rentevoet i . Indien men dus de vereenvoudigende veronderstelling maakt dat de aanvangssom alleen wordt gebruikt om de lijfrentes te betalen, dan zou de prijs moeten zijn

$$R_k = \frac{1}{P_k} \left(\frac{P_{k+1}}{1+i} + \frac{P_{k+2}}{(1+i)^2} + \frac{P_{k+3}}{(1+i)^3} + \dots \right). \quad (2.2)$$

Halley heeft op die manier waarden R_k verkregen, waarin R_k staat voor de factor waarmee de gewenste lijfrente moet worden vermenigvuldigd om de nodige aanvangssom te verkrijgen. Een man van 20 jaar zou dus elk jaar $1/12,78\%$ van de beginsom krijgen. Een man van 50 zou echter $1/9,21 \approx 10,9\%$ krijgen, omdat hij minder jaren te leven zou hebben. Het dubbele van de rente komt overeen met een lijfrente gelijk aan 12% van de oorspronkelijke som, of met een prijs gelijk aan $8,33$ maal de lijfrente.

Tabel 2.2: Vermenigvuldigingsfactor die de prijs van lijfrentes op leven geeft.

k	R_k	k	R_k	k	R_k	k	R_k	k	R_k
1	10,28	15	13,33	30	11,72	45	9,91	60	7,60
5	13,40	20	12,78	35	11,12	50	9,21	65	6,54
10	13,44	25	12,27	40	10,57	55	8,51	70	5,32

De berekeningen zijn natuurlijk nogal omslachtig. Halley kon echter gebruik maken van logaritmetabellen om de algemene term $P_{k+n}/(1+i)^n$ sneller te verkrijgen. Aangezien hij geen waarden voor P_k boven 84 jaar heeft gegeven, is het niet mogelijk zijn berekeningen precies te controleren. Tenslotte had het werk van Halley geen onmiddellijke gevolgen: gedurende verschillende decennia bleven lijfrentes op levens in Engeland en elders verkocht worden tegen een prijs die onafhankelijk was van de leeftijd van de koper en tegen een prijs die veel lager was dan hij zou kunnen zijn, bijvoorbeeld 7 maal de lijfrente.

De vragen die voortvloeiden uit sterftetafels interesseerden vele wetenschappers in Halleys tijd. De Nederlandse Christiaan Huygens, in 1657 auteur van het eerste aan de waarschijnlijkheidsleer gewijde manuscript, besprak in 1669 in zijn briefwisseling met zijn broer de sterftetabel van Graunt en de berekening van de levensverwachting².

²De levensverwachting op leeftijd k wordt verkregen uit formule (2.2) door $i = 0$ te nemen.

Enkele jaren voordat hij Neumann in contact bracht met de *Royal Society*, schreef ook Leibniz over de berekening van de levensverwachting in een essay dat ongepubliceerd bleef. In 1709 was het de beurt aan Nikolaus I Bernoulli. In 1725 publiceerde Abraham de Moivre een hele *Verhandeling over lijfrentes*. Hij merkte in het bijzonder op dat de factor R_k gemakkelijk kon worden berekend voor oude leeftijden omdat de formule (2.2) slechts een paar termen bevatte. Men kon dan de achterwaartse recurrentieformule gebruiken

$$R_k = \frac{P_{k+1}}{P_k} \frac{1 + R_{k+1}}{1 + i},$$

wat gemakkelijk te bewijzen is uitgaande van (2.2). Met behulp van de waarde die Halley geeft voor de factor op 70-jarige leeftijd, kan men dus de andere waarden van tabel 2.2 controleren³.

Na dit demografische intermezzo keerde Halley terug naar zijn belangrijkste onderzoeksonderwerpen. Tussen 1698 en 1700 zeilde hij over de Atlantische Oceaan om een kaart van het magnetisch veld van de aarde te tekenen. In 1704 werd hij professor aan de Universiteit van Oxford. Het jaar daarop publiceerde hij een boek over kometen en voorspelde dat de komeet van 1682, die Kepler in 1607 had waargenomen, in 1758 zou terugkomen: deze werd bekend als „Halley’s komeet”. Hij publiceerde ook een vertaling van het boek van Apollonius van Perga over kegelsneden. In 1720 verving hij Flamsteed als *Astronomer Royal*. Hij probeerde het probleem van het nauwkeurig bepalen van de lengtegraad op zee door waarneming van de maan op te lossen, een probleem van groot praktisch belang voor de navigatie. Hij overleed in 1742 in Greenwich op 86-jarige leeftijd.

Verdere lectuur

1. Fox, M.V.: *Scheduling the Heavens*. Morgan Reynolds (2007)
2. Graunt, J.: *Natural and Political Observations Mentioned in a Following Index and Made upon the Bills of Mortality* (1665). echo.mpiwg-berlin.mpg.de
3. Hald, A.: *A History of Probability and Statistics and Their Applications before 1750*. Wiley (2003).
4. Halley, E.: An estimate of the degrees of the mortality of mankind. *Phil. Trans. Roy. Soc. London* 17, 596–610 (1693). gallica.bnf.fr
5. Heyde, C.C.: John Graunt. In: Heyde, C.C., Seneta, E. (eds.) *Statisticians of the Centuries*, 14–16. Springer (2001)
6. Koch, P.: Caspar Neumann. In: *ibid.*, 29–32.
7. Le Bras, H.: *Naissance de la mortalité*. Gallimard, Paris (2000)

³Het lijkt erop dat er een paar fouten in de tabel staan, met name voor de leeftijden 5 en 15.

Hoofdstuk 3

Euler en de geometrische groei van populaties (1748–1761)

Euler schreef verschillende malen over populatiedynamica. In zijn verhandeling uit 1748, *Inleiding tot de analyse van het oneindige*, bevatte het hoofdstuk over de exponentiële functie vier voorbeelden over de exponentiële groei van een populatie. In 1760 publiceerde hij een artikel waarin hij deze exponentiële groei combineerde met een leeftijdsstructuur van de populatie. Dit werk is een voorloper van de theorie van de „stabiele” populatie, die in de twintigste eeuw werd ontwikkeld en een belangrijke rol speelt in de demografie. In 1761 hielp Euler ook Sommilch met de tweede editie van zijn verhandeling over demografie. Hij werkte een interessant model uit dat een variant is van de rij van Fibonacci, maar publiceerde zijn gedetailleerde analyse niet.

Leonhard Euler werd in 1707 geboren in Basel, Zwitserland. Zijn vader was een protestantse dominee. In 1720 begon Euler zijn studie aan de universiteit. Hij kreeg ook privé-lessen in wiskunde van Johann Bernoulli, een van de beroemdste wiskundigen van de generatie na Leibniz en Newton. Hij raakte bevriend met twee van Johann Bernoulli's zonen: Nikolaus II en Daniel. In 1727 voegde Euler zich bij Daniel aan de pas opgerichte Academie van Wetenschappen in Sint-Petersburg. Behalve in wiskunde was hij ook geïnteresseerd in natuurkunde en vele andere wetenschappelijke en technische onderwerpen. In 1741 nodigde koning Frederik II van Pruisen hem uit om directeur te worden van de afdeling Wiskunde van de Academie van Wetenschappen in Berlijn. Euler publiceerde een aanzienlijk aantal artikelen en boeken over vele aspecten van de mechanica (astronomie, elasticiteit, vloeistoffen, vaste stoffen) en de wiskunde (getaltheorie, algebra, oneindige reeksen, elementaire functies, complexe getallen, differentiaal- en integraalrekening, differentiaal- en partiële differentiaalvergelijkingen, optimalisatie, meetkunde), maar ook over demografie. Hij was de meest productieve wiskundige van zijn tijd.

In 1748 publiceerde Euler een verhandeling in het Latijn, getiteld *Introductio in analysin infinitorum* („Inleiding tot de analyse van het oneindige”). In het hoofdstuk over exponentiële en logaritmische functies behandelde hij



Figuur 3.1:
Euler (1707–1783)

zes voorbeelden: een over de wiskundige theorie van muziekschalen, een ander over de terugbetaling van een lening met rente, en vier over de dynamica van een populatie. In het laatste voorbeeld veronderstelde Euler dat de populatie P_n in jaar n voldoet aan

$$P_{n+1} = (1 + x)P_n$$

voor alle gehele getallen n . De groeisnelheid x is een positief reëel getal. Uitgaande van een beginvoorwaarde P_0 wordt de populatie in jaar n gegeven door

$$P_n = (1 + x)^n P_0.$$

Dit wordt geometrische of exponentiële groei genoemd. Het eerste voorbeeld vraagt:

„Als de populatie in een bepaalde streek jaarlijks met een dertigste toeneemt en er op een bepaald moment 100.000 inwoners zijn, zouden wij de populatiegrootte na 100 jaar willen kennen.”

Het antwoord is

$$P_{100} = (1 + 1/30)^{100} \times 100.000 \approx 2.654.874.$$

Voor dit voorbeeld liet Euler zich inspireren door de volkstelling van Berlijn die in 1747 had plaatsgevonden en die een schatting van de populatie opleverde van 107.224. Zijn berekening toont aan dat een populatie binnen een eeuw meer dan vertienvoudig kan zijn. Dit is precies wat men destijds voor de stad Londen had waargenomen.

Merk op dat het berekenen van $(1 + 1/30)^{100}$ heel gemakkelijk is met een moderne zakrekenmachine. In de tijd van Euler moest men echter logaritmen gebruiken om talrijke vermenigvuldigingen met de hand te vermijden en het resultaat snel te krijgen. Men berekent eerst de decimale logaritme (met basis 10) van P_{100} . Uit de fundamentele eigenschap van de logaritme $\log(ab) = \log a + \log b$ volgt dat

$$\log P_{100} = 100 \log(31/30) + \log(100.000) = 100(\log 31 - \log 30) + 5.$$

Logaritmen waren in 1614 geïntroduceerd door de Schot John Napier. Zijn vriend Henry Briggs had in 1617 de eerste tabel met decimale logaritmen gepubliceerd. In 1628 had de Nederlander Adriaan Vlacq het werk van Briggs voltooid door een tabel te publiceren met de decimale logaritmen van de gehele getallen van 1 tot 100.000, met een precisie van tien cijfers. Dit is het soort tabel dat Euler gebruikte om $\log 30 \approx 1,477121255$, $\log 31 \approx 1,491361694$, en tenslotte $\log P_{100} \approx 6,4240439$ te vinden. Nu moeten we nog het getal P_{100} vinden waarvan we de logaritme kennen. Aangezien de decimale logaritmen van gehele getallen van 1 tot 100.000 tussen 0 en 5 liggen, zoekt men in plaats daarvan de logaritme van $P_{100}/100$, die 4,4240439 bedraagt. Men kan in de tabel van logaritmen nagaan dat

$$\log 26.548 \approx 4,424031809 \quad \text{en} \quad \log 26.549 \approx 4,424048168.$$

Door de logaritmische functie te vervangen door een rechte lijn tussen 26.548 en 26.549 verkrijgt Euler dat

$$\frac{P_{100}}{100} \approx 26.548 + \frac{4,4240439 - 4,424031809}{4,424048168 - 4,424031809} \approx 26.548,74.$$

Dus $P_{100} \approx 2.654.874$.

Het tweede voorbeeld van populatiedynamica in Eulers boek is het volgende:

„Aangezien na de zondvloed alle mensen afstammen van een populatie van zes, zouden wij, veronderstellende dat de populatie na tweehonderd jaar 1.000.000 was, het jaarlijkse groeipercen- tage willen vinden.”

Aangezien

$$10^6 = (1+x)^{200} \times 6,$$

krijgen we met een zakrekenmachine $x = (10^6/6)^{1/200} - 1 \approx 0,061963$. Met de tafels van logaritmen moet men $\log(10^6) = 200 \log(1+x) + \log 6$ doorlopen om $\log(1+x) = (6 - \log 6)/200$ te krijgen en $1+x \approx 1,061963$. Zo

kon Euler concluderen dat de populatie zou toenemen met $x \approx 1/16$ per jaar. Om de oorsprong van dit voorbeeld te begrijpen, moet men bedenken dat de filosofen van die tijd de waarheid van de bijbelse verhalen begonnen te ontkennen. Een letterlijke lezing zou het tijdstip van de zondvloed vaststellen rond 2350 v. Chr. met de volgende overlevenden: Noach, zijn drie zonen en hun vrouwen. Het boek Genesis zegt:

„Deze drie waren de zonen van Noach; en van dezen is de ganse aarde overspreid.”

Een populatiegroei van $1/16$ (of 6,25 %) per jaar na de zondvloed leek Euler niet al te onrealistisch. Als zoon van een protestantse dominee en zijn hele leven gelovig gebleven, concludeerde hij:

„Daarom is het belachelijk voor de ongelovigen om te beweren dat in zo'n korte tijdspanne de hele aarde niet bevolkt kan zijn, beginnend met één enkele man.”¹

Euler merkte ook op dat als de groei in hetzelfde tempo was doorgegaan tot 400 jaar na de zondvloed, de populatie $(1+x)^{400} \times 6 = (10^6/6)^2 \times 6 = 166$ miljard zou zijn geweest:

„Maar de hele aarde zou nooit in staat zijn om die populatie te onderhouden.”

Dit idee zou een halve eeuw later door Malthus verder worden uitgewerkt (zie hoofdstuk 5).

Euler's derde voorbeeld vraagt:

„Als de menselijke populatie elke eeuw verdubbelt, wat is dan het jaarlijkse groeipercentage?”

Aangezien $(1+x)^{100} = 2$, krijgen we met een zakrekenmachine $x = 2^{1/100} - 1 \approx 0,00695$. Met de tafels van logaritmen, $100 \log(1+x) = \log 2$. Dus $\log(1+x) \approx 0,0030103$ en $1+x \approx 1,00695$. De populatie groeit dus elk jaar met $x \approx 1/144$. Het vierde en laatste voorbeeld vraagt op dezelfde manier:

¹In het boek dat Graunt in 1662 publiceerde (zie hoofdstuk 2), vindt men een soortgelijke opmerking:

„Eén paar, Adam en Eva, dat zich elke 64 jaar verdubbelt gedurende 5.160 jaar, de leeftijd van de wereld volgens het Schrift, zal veel meer mensen voortbrengen dan er nu op de wereld zijn. Daarom is de wereld niet meer dan 100 duizend jaar oud, zoals sommigen zich ijdel voorstellen, noch meer dan wat het Schrift ervan maakt.”

„Als de menselijke populatie jaarlijks met $1/100$ toeneemt, willen wij graag weten hoe lang het zal duren voordat de populatie tien keer zo groot is geworden.”

Met $(1 + 1/100)^n = 10$, vinden we $n \log(101/100) = 1$. Dus $n = 1/(\log 101 - 2) \approx 231$ jaar. Dit is alles wat er in de Inleiding tot de Analyse van het On-eindige uit 1748 te vinden is over de dynamica van een populatie. Euler zou enkele jaren later dieper op dit onderwerp ingaan.

In 1760 publiceerde hij in de handelingen van de Akademie der Wissenschaften in Berlijn een werk getiteld *Een algemeen onderzoek naar de sterfte en vermeerdering van de menselijke soort*. Dit werk was een soort synthese tussen zijn eerdere analyse van de geometrische groei van de populatie en eerdere studies over sterftetabellen (zie hoofdstuk 2). Euler beschouwde bijvoorbeeld het probleem:

„Het kennen van het aantal geboorten en begrafenissen in de loop van een jaar, om het aantal van alle levenden te vinden en hun jaarlijkse toename, voor een gegeven hypothese van mortaliteit.”

Euler veronderstelde hier dat de volgende getallen bekend zijn:

- het aantal geboorten B_n in het jaar n ;
- het aantal sterfgevallen D_n in het jaar n ;
- het aandeel q_k van pasgeborenen die de leeftijd van $k \geq 1$ bereiken.

Stel P_n is de populatie in jaar n . Euler maakte twee extra impliciete veronderstellingen:

- de populatie neemt geometrisch toe: $P_{n+1} = rP_n$ (met $r = 1 + x$);
- de verhouding tussen geboorten en populatie is constant: $B_n/P_n = m$.

Deze twee aannames impliceren dat het aantal geboorten geometrisch en aan hetzelfde tempo toeneemt: $B_{n+1} = rB_n$. Euler beschouwde vervolgens de toestand van de populatie in het interval van honderd jaar, zeg tussen de jaren $n = 0$ en $n = 100$, in de veronderstelling dat niemand langer dan honderd jaar overleeft. Om deze voorstelling te verduidelijken, noemen we $P_{k,n}(k, 1)$ de populatie geboren in het jaar $n - k$, die leeft aan het begin van het jaar n . Noem $P_{0,n} = B_n$ het aantal geboorten in het jaar n . Uit de definitie van de overlevingscoëfficiënt q_k volgt $P_{k,n} = q_k P_{0,n-k} = q_k B_{n-k}$. Dus

$$\begin{aligned} r^{100} P_0 = P_{100} &= P_{0,100} + P_{1,100} + \cdots + P_{100,100} \\ &= B_{100} + q_1 B_{99} + \cdots + q_{100} B_0 \\ &= (r^{100} + r^{99} q_1 + \cdots + q_{100}) B_0. \end{aligned}$$

Als we deze vergelijking delen door $r^{100} P_0$, krijgen we

$$1 = m \left(1 + \frac{q_1}{r} + \frac{q_2}{r^2} + \cdots + \frac{q_{100}}{r^{100}} \right). \quad (3.1)$$

Dit is de vergelijking die in de demografie soms de „vergelijking van Euler” wordt genoemd. Als we geboorten en sterfgevallen apart tellen, krijgen we

$$r P_n = P_{n+1} = P_n - D_n + B_{n+1} = P_n - D_n + r B_n. \quad (3.2)$$

Het aantal sterfgevallen neemt dus ook geometrisch toe: $D_{n+1} = r D_n$. Bovendien,

$$\frac{1}{m} = \frac{P_n}{B_n} = \frac{D_n/B_n - r}{1 - r}. \quad (3.3)$$

Als we dit vervangen door vergelijking (3.1), komen we uiteindelijk uit op de vergelijking

$$\frac{D_n/B_n - 1}{1 - r} = \frac{q_1}{r} + \frac{q_2}{r^2} + \cdots + \frac{q_{100}}{r^{100}}, \quad (3.4)$$

waarbij er slechts één onbekende overblijft: r . Dit is wat men gewoonlijk een impliciete vergelijking noemt, omdat men r niet kan extraheren als functie van de andere parameters. Maar we kunnen wel het linker- en rechterdeel van vergelijking (3.4) berekenen voor een vaste waarde van r en r laten variëren tot de twee zijden gelijk zijn. De aldus verkregen waarde van r geeft de groeisnelheid $x = r - 1$ van de populatie. Merk op dat uit vergelijkingen (3.1) en (3.3) voor de populatie P_n de volgende uitdrukking volgt:

$$P_n = B_n \left(1 + \frac{q_1}{r} + \frac{q_2}{r^2} + \cdots + \frac{q_{100}}{r^{100}} \right).$$

Wanneer de populatie stationair is ($r = 1$), is deze uitdrukking dezelfde als die welke door Halley werd gebruikt om de populatie van de stad Breslau te schatten (zie hoofdstuk 2).

Euler overwoog ook de volgende vraag:

„Met de hypothesen van sterfte en vruchtbaarheid kan, als men het aantal levenden kent, worden berekend hoeveel er op elke leeftijd zullen zijn.”

Aangezien de overlevingscoëfficiënten q_k en de vruchtbaarheidscoëfficiënt m bekend zijn, kan het groeipercentage r worden berekend uit vergelijking (3.1). In het jaar n is het aantal mensen dat in het jaar $n - k$ geboren wordt $q_k B_{n-k} = q_k B_n / r^k$ (met $q_0 = 1$). Dus het deel van de totale populatie met leeftijd k is

$$\frac{q_k / r^k}{1 + q_1 / r + q_2 / r^2 + \cdots + q_{100} / r^{100}}.$$

Deze verhouding is constant. In de terminologie van Lotka (zie hoofdstuk 10) wordt gezegd dat de populatie „stabil” is: de leeftijds piramide behoudt dezelfde vorm doorheen de tijd.

Euler onderzocht vervolgens opnieuw het probleem van het opstellen van een sterftetabel wanneer de populatie niet stationair is maar geometrisch groeit:

„Als men het aantal levenden kent, vergelijkt men het aantal geboorten met het aantal sterfgevallen op elke leeftijd in de loop van een jaar, om de wet van de sterfte te vinden.”

Met „mortaliteitswet” bedoelde Euler de verzameling overlevingscoëfficiënten q_k . De totale populatie wordt nu verondersteld bekend te zijn door een volkstelling, wat bij Halley niet het geval was (zie hoofdstuk (2)). Uit vergelijking (3.2) blijkt de groeisnelheid

$$r = \frac{P_n - D_n}{P_n - B_n}.$$

Zij $D_{k,n}$ het aantal personen dat in het jaar n op leeftijd k sterft: deze personen zijn geboren in het jaar $n - k$. Dus $D_{k,n} = (q_k - q_{k+1})B_{n-k}$. Maar $B_{n-k} = B_n/r^k$. De overlevingscoëfficiënten q_k kunnen dus berekend worden met de recurrente betrekking

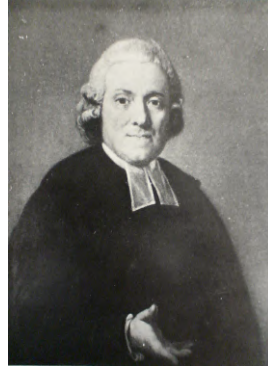
$$q_{k+1} = q_k - \frac{r^k D_{k,n}}{B_n}$$

voor alle $k \geq 0$, met $q_0 = 1$. Vermenigvuldigd met B_n geeft deze vergelijking terug de formule (2.1) gebruikt door Halley voor het stationaire geval $r = 1$. Euler benadrukte niettemin dat zijn methode om de overlevingscoëfficiënten q_k te berekenen, veronderstelt dat de populatie stelselmatig toeneemt, zonder verstoringen zoals pestepidemieën, oorlogen, hongersnoden enz. Indien de volkstellingen in Eulers tijd de leeftijd van de mensen hadden geregistreerd (zoals in Zweden), zou deze veronderstelling overbodig zijn geweest en hadden de coëfficiënten q_k gemakkelijker kunnen worden berekend.

Gegeven de overlevingscoëfficiënten q_k , toonde Euler ook hoe men de prijs van lijfrenten op leven berekent. Hierbij maakte hij geen melding van de werken van Halley of de Moivre over dit onderwerp. Euler gebruikte een rentevoet van 5 % en de sterftetabel die in 1742 door de Nederlander Willem Kersseboom werd gepubliceerd.

Euler was niet de enige wetenschapper aan de Berlijnse Academie die zich met demografie bezighield. Zijn collega Johann Peter Süssmilch publiceerde in 1741 een verhandeling in het Duits, getiteld *De Goddelijke orde zichtbaar in de geboorten, het sterven en de voortplanting van het menselijk*

geslacht, die tegenwoordig wordt beschouwd als de eerste volledig aan demografie gewijde verhandeling. Süßmilch schreef in 1752 ook een boek getiteld *Over de snelle groei van de stad Berlijn*.



Figuur 3.2:
Süßmilch (1707–1767)

In 1761 publiceerde Süßmilch een tweede editie van zijn verhandeling. In het hoofdstuk getiteld *Over de snelheid van toename en over de verdubbelingstijd van populaties*, nam hij een interessant wiskundig model op dat Euler voor hem had uitgewerkt. Het model was vergelijkbaar met dat van Fibonacci (zie hoofdstuk 1), maar dan voor een menselijke populatie. Uitgaande van een koppel (een man en een vrouw) die beiden 20 jaar oud zijn in het jaar 0, veronderstelde Euler dat mensen op 40-jarige leeftijd sterven en op 20-jarige leeftijd trouwen, terwijl elk koppel zes kinderen krijgt: twee kinderen (een jongen en een meisje) op 22-jarige leeftijd, nog eens twee op 24-jarige leeftijd en de laatste twee op 26-jarige leeftijd. Door de jaren twee aan twee te tellen, zodat B_i het aantal geboorten in het jaar $2i$ is, concludeert Euler dat

$$B_i = B_{i-11} + B_{i-12} + B_{i-13} \quad (3.5)$$

voor alle $i \geq 1$. De beginvoorwaarden komen overeen met $B_{-12} = 0$, $B_{-11} = 0$, $B_{-10} = 2$ en $B_i = 0$ voor $-9 \leq i \leq 0$. Euler kon zo het aantal geboorten berekenen, zoals weergegeven in de tweede kolom van tabel 3.1. Het aantal sterfgevallen D_i in jaar $2i$ is dan gelijk aan het aantal geboorten in jaar $2i - 40$: $D_i = B_{i-20}$ voor $i \geq 10$ terwijl $D_i = 0$ voor $i \leq 9$. Het aantal mensen dat in jaar $2i$ in leven is, P_i , is gelijk aan het aantal mensen dat in jaar $2i - 2$ in leven is, plus het aantal geboorten in jaar $2i$, min het aantal sterfgevallen in jaar $2i$: $P_i = P_{i-1} + B_i - D_i$.

Dit hoofdstuk in het boek van Süßmilch eindigt met een opmerking die al gemaakt had kunnen zijn over de reeks van Fibonacci:

Tabel 3.1: De tafel van Euler.

i	Geboorten	Sterfgevallen	Leven
0	0	0	2
1	2	0	4
2	2	0	6
3	2	0	8
4	0	0	8
5	0	0	8
6	0	0	8
7	0	0	8
8	0	0	8
9	0	0	8
10	0	2	6
11	0	0	6
12	2	0	8
13	4	0	12
14	6	0	18
15	4	0	22
16	2	0	24
17	0	0	24
18	0	0	24
19	0	0	24
20	0	0	24
21	0	2	22
22	0	2	20
23	2	2	20
24	6	0	26
25	12	0	38
26	14	0	52
27	12	0	64
28	6	0	70
29	2	0	72
30	0	0	72
31	0	0	72
32	0	2	70
33	0	4	66
34	2	6	62
35	8	4	66
36	20	2	84
37	32	0	116
38	38	0	154
39	32	0	186

i	Geboorten	Sterfgevallen	Leven
40	20	0	206
41	8	0	214
42	2	0	216
43	0	2	214
44	0	6	208
45	2	12	198
46	10	14	194
47	30	12	212
48	60	6	266
49	90	2	354
50	102	0	456
51	90	0	546
52	60	0	606
53	30	0	636
54	10	2	644
55	2	8	638
56	2	20	620
57	12	32	600
58	42	38	604
59	100	32	672
60	180	20	832
61	252	8	1.076
62	282	2	1.356
63	252	0	1.608
64	180	0	1.788
65	100	2	1.886
66	42	10	1.918
67	14	30	1.902
68	16	60	1.858
69	56	90	1.824
70	154	102	1.876
71	322	90	2.108
72	532	60	2.580
73	714	30	3.264
74	786	10	4.040
75	714	2	4.752
76	532	2	5.282
77	322	12	5.592
78	156	42	5.706
79	72	100	5.678

i	Geboorten	Sterfgevallen	Leven
80	86	180	5.584
81	226	252	5.558
82	532	282	5.808
83	1.008	252	6.564
84	1.568	180	7.952
85	2.032	100	9.884
86	2.214	42	12.056
87	2.032	14	14.074
88	1.568	16	15.626
89	1.010	56	16.580
90	550	154	16.976
91	314	322	16.968
92	384	532	16.820
93	844	714	16.950
94	1.766	786	17.930
95	3.108	714	20.324
96	4.608	532	24.400
97	5.814	322	29.892
98	6.278	156	36.014
99	5.814	72	41.756
100	4.610	86	46.280
101	3.128	226	49.182
102	1.874	532	50.524
103	1.248	1.008	50.764
104	1.542	1.568	50.738
105	2.994	2.032	51.700
106	5.718	2.214	55.204
107	9.482	2.032	62.654
108	13.530	1.568	74.616
109	16.700	1.010	90.306
110	17.906	550	107.662
111	16.702	314	124.050
112	13.552	384	137.218
113	9.612	844	145.986
114	6.250	1.766	150.470
115	4.664	3.108	152.026
116	5.784	4.608	153.202
117	10.254	5.814	157.642
118	18.194	6.278	169.558
119	28.730	5.814	192.474

„De grote wanorde die in de tabel van Euler schijnt te heersen verhindert niet dat het aantal geboorten een soort progressie volgt die men terugkerende reeksen noemt [...] Wat ook de aanvankelijke wanorde van deze progressies is, zij veranderen in een geometrische progressie als zij niet worden onderbroken en de wanorde van het begin vervaagt beetje bij beetje en verdwijnt bijna geheel.”

Het boek zegt niet meer over de wiskunde van dit populatiemodel. Euler ging echter veel verder in een manuscript getiteld *Over de vermenigvuldiging van het menselijk ras*, dat tijdens zijn leven ongepubliceerd bleef. Op zoek naar een oplossing van vergelijking (3.5) van de vorm $B_i = c r^i$, d.w.z. van de vorm van een geometrische progressie, verkreeg hij na vereenvoudiging een polynomiale vergelijking van graad 13:

$$r^{13} = r^2 + r + 1. \quad (3.6)$$

Hij zocht een oplossing in de buurt van $r = 1$ en merkte, met behulp van een tabel met logaritmen voor de berekening van r^{13} , dat

$$1 + r + r^2 - r^{13} \approx \begin{cases} 0,212 & \text{als } r = 1,09, \\ -0,142 & \text{als } r = 1,10. \end{cases}$$

Dus vergelijking (3.6) heeft een wortel tussen 1,09 en 1,10. Door de functie $1 + r + r^2 - r^{13}$ te benaderen met een lijnstuk op dit interval, verkrijgt Euler

$$r \approx \frac{0,142 \times 1,09 + 0,212 \times 1,10}{0,142 + 0,212} \approx 1,0960.$$

Aangezien de jaren twee aan twee worden geteld, wordt het aantal geboorten elk jaar met \sqrt{r} vermenigvuldigd. Dit aantal verdubbelt elke n jaar als $(\sqrt{r})^n = 2$, d.w.z. elke $n = 2 \log 2 / \log r \approx 15$ jaar. Aangezien asymptotisch $B_i \approx c r^i$ en aangezien het aantal sterfgevallen D_i in jaar $2i$ gelijk is aan B_{i-20} , krijgen we $D_i \approx B_i / r^{20}$ met $r^{20} \approx 6,25$. Het aantal geboorten is ongeveer zes keer zo groot als het aantal sterfgevallen. Aangezien het aantal levenden in het jaar $2i$ gelijk is aan $B_i + B_{i-1} + \dots + B_{i-19}$, bekomen we ook dat

$$P_i \approx B_i \left(1 + \frac{1}{r} + \dots + \frac{1}{r^{19}} \right) = B_i \frac{1 - r^{-20}}{r^{19} - r^{20}} \approx 9,59 B_i.$$

De totale populatie is ongeveer tien keer zo groot als het aantal geboorten.

Het bewijs dat de reeks (B_i) in tabel 3.1 inderdaad asymptotisch groeit als r^i is gecompliceerder. Sinds het werk van Abraham de Moivre over recurrente betrekkingen was bekend dat, door de invoering van de „genererende functie”

$$f(x) = \sum_{i=0}^{+\infty} B_i x^i,$$

kan men $f(x)$ uitdrukken als een rationale functie. Euler had de methode uitgelegd in zijn *Inleiding tot de Analyse van het Oneindige* in 1748: de recurrente betrekking (3.5) geeft inderdaad

$$\begin{aligned} f(x) &= \sum_{i=0}^{12} B_i x^i + \sum_{i=13}^{+\infty} (B_{i-11} + B_{i-12} + B_{i-13}) x^i \\ &= 2x + 2x^2 + 2x^3 + 2x^{12} + f(x) (x^{11} + x^{12} + x^{13}). \end{aligned}$$

Dus

$$f(x) = \frac{2x + 2x^2 + 2x^3 + 2x^{12}}{1 - x^{11} - x^{12} - x^{13}}.$$

Euler wist dat zo'n rationale functie ontleed kon worden in de vorm

$$f(x) = \frac{a_1}{1 - \frac{x}{x_1}} + \dots + \frac{a_{13}}{1 - \frac{x}{x_{13}}},$$

waarbij de getallen x_1, \dots, x_{13} de reële of complexe wortels zijn van de vergelijking $1 - x^{11} - x^{12} - x^{13} = 0$. Dus

$$f(x) = \sum_{i \geq 0} a_1 \left(\frac{x}{x_1} \right)^i + \dots + a_{13} \left(\frac{x}{x_{13}} \right)^i.$$

Aangezien B_i de coëfficiënt is van x^i in $f(x)$, bewam Euler dat

$$B_i = \frac{a_1}{(x_1)^i} + \dots + \frac{a_{13}}{(x_{13})^i} \approx \frac{a_k}{(x_k)^i}$$

als $i \rightarrow +\infty$, waarbij x_k de wortel met de kleinste modulus is. Met andere woorden, B_i neigt naar een geometrische groei als $(1/x_k)^i$. Restte nog op te merken dat x_k een wortel is van de vergelijking $1 - x^{11} - x^{12} - x^{13} = 0$ als en slechts als $r = 1/x_k$ een wortel is van vergelijking (3.6). Bepaalde details van het bewijs werden uiteindelijk opgehelderd door Gumbel in 1916.

In 1765 publiceerde Süßmilch een derde druk van zijn verhandeling en in 1767 overleed hij in Berlijn. Op slechte voet met de koning van Pruisen keerde Euler in 1766 terug naar Sint-Petersburg. Ondanks het verlies van zijn gezichtsvermogen bleef hij met de hulp van zijn zonen en collega's een groot aantal werken publiceren, vooral over algebra, integraalrekening, optica en scheepsbouw. Zijn *Brieven gericht aan een Duitse prinses over verschillende onderwerpen in de natuurlijke filosofie*, geschreven in Berlijn tussen 1760 en 1762, werd gepubliceerd tussen 1768 en 1772 en werd een bestseller in heel Europa. Euler stierf in Sint-Petersburg in 1783. Zijn bijdrage aan de mathematische demografie, vooral zijn analyse van de „stabiele” leeftijds piramide in een exponentieel groeiende populatie, zou pas in de twintigste eeuw herontdekt worden (zie de hoofdstukken 10 en 21).

Verdere lectuur

1. Euler, L.: Recherches générales sur la mortalité et la multiplication du genre humain. *Hist. Acad. R. Sci. B.-Lett. Berl.* 16, 144–164 (1760). eulerarchive
2. Euler, L.: Sur la multiplication du genre humain. In: *Leonhardi Euleri Opera omnia*, Ser. I, vol. 7, 545–552. Teubner, Leipzig (1923)
3. Euler, L.: *Introductio in analysin infinitorum* (1748). → *Leonhardi Euleri Opera omnia*, Ser. I, vol. 8, Teubner, Leipzig (1922). gallica.bnf.fr
4. Fellmann, E.A.: *Leonhard Euler*. Birkhäuser, Basel (2007)
5. Gumbel, E.J.: Eine Darstellung statistischer Reihen durch Euler. *Jahresber. dtsh. Math. Ver.* 25, 251–264 (1917). digizeitschriften.de
6. Reimer, K.F.: Johann Peter Süßmilch, seine Abstammung und Biographie. *Arch. soz. Hyg. Demogr.* 7, 20–28 (1932)
7. Rohrbasser, J.M.: Johann Peter Süßmilch. In: Heyde, C.C., Seneta, E. (eds.) *Statisticians of the Centuries*, 72–76. Springer (2001)
8. Süßmilch, J.P.: *Die göttliche Ordnung*. Berlin (1761). mpiwg-berlin.mpg.de
9. Warusfel, A.: *Euler, les mathématiques et la vie*. Vuibert, Paris (2009)

Hoofdstuk 4

Daniel Bernoulli, d'Alembert en de inenting van pokken (1760)

In 1760 schreef Daniel Bernoulli een artikel waarin hij de pokken modelleerde. In zijn tijd was er veel controverse rond inenting, een praktijk die mensen kon beschermen maar ook dodelijk kon zijn. Hij gebruikte de levenstabel van Halley en enkele gegevens over pokken om aan te tonen dat inenting voordelig was als de kans om te sterven kleiner was dan 11 %. Inenting zou de levensverwachting bij de geboorte tot drie jaar kunnen verhogen. D'Alembert bekritiseerde het werk van Bernoulli, dat het eerste wiskundige model in de epidemiologie was.

Daniel Bernoulli werd in 1700 geboren in Groningen, Nederland. Zijn familie telde al twee beroemde wiskundigen: zijn vader Johann Bernoulli en zijn oom Jakob Bernoulli. In 1705 verhuisde Johann naar Basel in Zwitserland waar hij het professoraat aanvaardde dat was vrijgekomen door de dood van Jakob. Johann wilde niet dat zijn zoon wiskunde ging studeren. Daniel ging geneeskunde studeren en promoveerde in 1721 op een proefschrift over de ademhaling. Hij verhuisde naar Venetië en begon zich toe te leggen op de wiskunde, en publiceerde een boek in 1724. Nadat hij datzelfde jaar een prijs had gewonnen van de Parijse Academie van Wetenschappen voor een essay *Over de volmaaktheid van de zandloper op een schip op zee*, kreeg hij een hoogleraarschap aan de nieuwe Academie van Sint Petersburg. In deze jaren werkte hij vooral aan recurrente betrekkingen of aan de „paradox van Sint-Petersburg” in de waarschijnlijkheidsleer. In 1733 keerde Daniel Bernoulli terug naar de universiteit van Basel, waar hij achtereenvolgens plantkunde, fysiologie en natuurkunde doceerde. In 1738 publiceerde hij een boek over vloeistofdynamica dat beroemd is gebleven in de geschiedenis van de fysica. Rond 1753 raakte hij tegelijk met Euler en d'Alembert geïnteresseerd in het probleem van de trillende snaren, dat een belangrijke wiskundige controverse veroorzaakte.

In 1760 legde hij aan de Academie van Wetenschappen in Parijs een werk voor met de titel *Een poging tot een nieuwe analyse van de sterfte veroorzaakt door pokken en van de voordelen van inenting ter voorkoming daarvan*. De vraag was of inenting (het vrijwillig inbrengen van een kleine hoeveelheid



Figuur 4.1:
Daniel Bernoulli
(1700–1782)

minder virulente pokken in het lichaam om het te beschermen tegen latere infecties) moet worden aangemoedigd, ook al is het soms een dodelijke ingreep. Deze techniek was reeds lang bekend in Azië en was in 1718 in Engeland ingevoerd door Lady Montagu, de echtgenote van de Britse ambassadeur in het Ottomaanse Rijk. In Frankrijk werd, ondanks de dood van de oudste zoon van Lodewijk XIV aan de pokken in 1711, inenting met tegenzin overwogen. Voltaire, die in 1723 de pokken had overleefd en een aantal jaren in ballingschap in Engeland had geleefd om de nieuwste innovaties te observeren, pleitte in 1734 in zijn *Filosofische Brieven voor inenting*. De Franse wetenschapper La Condamine, die ook de pokken had overleefd, pleitte voor inenting op de Academie van Wetenschappen in Parijs in 1754.

Voordat hij in 1759 in Bazel overleed, moedigde Maupertuis Daniel Bernoulli aan om het inentingsprobleem vanuit wiskundig oogpunt te bestuderen. Meer bepaald ging het erom een manier te vinden om het voordeel op lange termijn van inenting te vergelijken met het onmiddellijke risico om te sterven. Voor dit doel maakte Bernoulli de volgende vereenvoudigende veronderstellingen:

- mensen die voor het eerst met pokken besmet raken, sterven met een waarschijnlijkheid p (onafhankelijk van de leeftijd) en overleven met een waarschijnlijkheid $1 - p$;
- iedereen heeft een kans q om elk jaar besmet te raken; meer precies is de kans voor één individu om tussen leeftijd x en leeftijd $x + dx$ besmet te raken $q dx$, waarbij dx een infinitesimale tijdsperiode is;
- mensen die de pokken overleven, zijn voor de rest van hun leven beschermd tegen nieuwe infecties (zij zijn immers ingeënt).

Zij $m(x)$ de sterfte op leeftijd x door andere oorzaken dan pokken: de kans

dat een individu sterft in een infinitesimaal tijdsbestek dx tussen leeftijd x en leeftijd $x + dx$ is $m(x)dx$. Beschouwen we een groep van P_0 mensen die in hetzelfde jaar geboren zijn, dan noemen we

- $S(x)$ het aantal vatbare mensen die op leeftijd x nog in leven zijn zonder ooit met pokken besmet te zijn geweest;
- $R(x)$ het aantal nog levende mensen van leeftijd x die pokken hebben overleefd;
- $P(x) = S(x) + R(x)$ het totale aantal nog levende mensen op leeftijd x .

Geboorte komt overeen met leeftijd $x = 0$. Dus $S(0) = P(0) = P_0$ en $R(0) = 0$. Door de berekeningsmethoden toe te passen die op het einde van de 17de eeuw door Newton, Leibniz en later door zijn vader ontwikkeld werden, stelde Daniel Bernoulli vast dat tussen de leeftijd x en de leeftijd $x + dx$ (met dx oneindig klein), elk vatbaar individu een kans qdx heeft om met pokken besmet te worden en een kans $m(x)dx$ om aan andere oorzaken te sterven. De variatie van het aantal vatbare personen is dus $dS = -Sdx - Sm(x)dx$, wat leidt tot de differentiaalvergelijking

$$\frac{dS}{dx} = -qS - m(x)S. \quad (4.1)$$

In deze vergelijking wordt dS/dx de afgeleide van de functie $S(x)$ genoemd. Gedurende hetzelfde kleine tijdsinterval is het aantal mensen dat aan pokken sterft $pSqdx$ en het aantal mensen dat pokken overleeft $(1-p)Sqdx$. Bovendien zijn er ook mensen die sterven aan andere oorzaken dan pokken. Dit leidt tot een tweede differentiaalvergelijking:

$$\frac{dR}{dx} = q(1-p)S - m(x)R. \quad (4.2)$$

Als we de twee vergelijkingen optellen, krijgen we

$$\frac{dP}{dx} = -pqS - m(x)P. \quad (4.3)$$

Uit vergelijkingen (4.1) en (4.3) kon Bernoulli aantonen dat de fractie van mensen die nog vatbaar zijn op leeftijd x is

$$\frac{S(x)}{P(x)} = \frac{1}{(1-p)e^{qx} + p}. \quad (4.4)$$

Om formule (4.4) te krijgen, elimineerde Bernoulli $m(x)$ uit de vergelijkingen (4.1) en (4.3):

$$-m(x) = q + \frac{1}{S} \frac{dS}{dx} = pq \frac{S}{P} + \frac{1}{P} \frac{dP}{dx}.$$

Na herschikking volgt

$$\frac{1}{P} \frac{dS}{dx} - \frac{S}{P^2} \frac{dP}{dx} = -q \frac{S}{P} + pq \left(\frac{S}{P} \right)^2.$$

We zien dat het linkerlid de afgeleide is van $f(x) = S(x)/P(x)$, dat is de fractie van vatbare mensen in de populatie van x . Dus

$$\frac{df}{dx} = -qf + pqf^2. \quad (4.5)$$

De oplossing van dit soort vergelijking was al tientallen jaren bekend dankzij het werk van Jakob Bernoulli, de oom van Daniel. Door de vergelijking te delen door f^2 en $g(x) = 1/f(x)$ te stellen, zien we dat $dg/dx = qg - pq$ en dat $g(0) = 1/f(0) = 1$. Stel $h(x) = g(x) - p$, dan krijgen we $dh/dx = qh$. Dus $h(x) = h(0)e^{qx} = (1-p)e^{qx}$. Tenslotte is $g(x) = (1-p)e^{qx} + p$ en $f(x) = 1/g(x)$.

Om zijn theorie toe te passen, gebruikte Bernoulli de sterftetabel van Halley (zie hoofdstuk 2). Deze tabel geeft het aantal mensen dat nog in leven is aan het begin van jaar x (met $x = 1, 2, \dots$) onder een cohort van 1.238 mensen geboren in jaar 0. Maar in het kader van zijn model had Bernoulli het aantal mensen $P(x)$ nodig dat daadwerkelijk de leeftijd x bereikt, en dat is iets anders. Omdat Bernoulli - zoals de meeste van zijn tijdgenoten - zich het verschil niet realiseerde (Halley's artikel is inderdaad niet erg expliciet), behield hij de getallen in Halley's tabel behalve het eerste getal 1.238, dat hij verving door 1.300 om een realistische sterfte tijdens het eerste levensjaar te krijgen. Deze getallen staan in de tweede kolom van tabel 4.1.

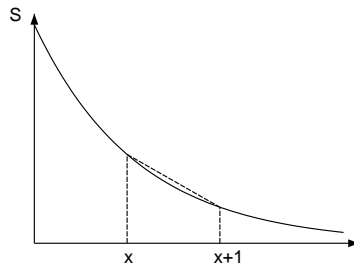
Bernoulli koos voor de kans om aan pokken te sterven $p = 1/8 = 12,5\%$, wat overeenkomt met de waarnemingen van zijn tijd. De jaarlijkse kans om pokken op te lopen q kon niet rechtstreeks worden geschat. Bernoulli heeft dus waarschijnlijk verschillende waarden voor q uitgetoet en uiteindelijk die waarde gekozen waarmee het aantal sterfgevallen ten gevolge van pokken na alle onderstaande berekeningen ongeveer 1/13 van het totale aantal sterfgevallen bedraagt, een verhouding die toen in verschillende Europese steden

Tabel 4.1: Halley's sterftetabel en Bernoulli's berekeningen.

Leeftijd x	Levend $P(x)$	Vatbaar $S(x)$	Immuun $R(x)$	Pokken sterfgevallen	Geen pokken $P^*(x)$
0	1.300	1.300	0	17,2	1.300
1	1.000	896	104	12,3	1.015
2	855	685	170	9,8	879
3	798	571	227	8,2	830
4	760	485	275	7,0	799
5	732	416	316	6,1	777
6	710	359	351	5,2	760
7	692	311	381	4,6	746
8	680	272	408	4,0	738
9	670	238	432	3,5	732
10	661	208	453	3,0	726
11	653	182	471	2,7	720
12	646	160	486	2,3	715
13	640	140	500	2,1	711
14	634	123	511	1,8	707
15	628	108	520	1,6	702
16	622	94	528	1,4	697
17	616	83	533	1,2	692
18	610	72	538	1,1	687
19	604	63	541	0,9	681
20	598	55	543	0,8	676
21	592	49	543	0,7	670
22	586	42	544	0,6	664
23	579	37	542	0,5	656
24	572	32	540		649
⋮	⋮	⋮	⋮	⋮	⋮

was waargenomen. De keuze $q = 1/8$ per jaar bleek een goede fit te geven, zoals we nu zullen zien¹.

Met formule (4.4) en de waarden van $P(x)$ in de tweede kolom van de tabel kunnen we het aantal $S(x)$ van de vatbare personen van x berekenen: dit is de derde kolom van de tabel, afgerond op het dichtstbijzijnde gehele getal. In de vierde kolom staat het aantal $R(x) = P(x) - S(x)$ van de personen van x die de pokken hebben overleefd. In de vijfde kolom staat in de rij die overeenkomt met leeftijd x het aantal sterfgevallen ten gevolge van pokken tussen leeftijd x en leeftijd $x + 1$. In theorie zou dit aantal de integraal $pq \int_x^{x+1} S(t) dt$ maar de formule $pq [S(x) + S(x+1)]/2$ geeft een goede benadering, zoals geschetst in figuur 4.2: de oppervlakte van het trapezium ligt dicht bij de oppervlakte onder de kromme, d.w.z. bij de integraal van de functie.



Figuur 4.2: De oppervlakte van het gestippelde trapezium benadert de integraal van de functie S tussen x en $x + 1$.

Bernoulli merkte op dat de som van alle getallen in de vijfde kolom 98 sterfgevallen aan pokken oplevert vóór de leeftijd van 24 jaar. Als we de tabel zouden voortzetten voor oudere leeftijden, zouden we nog maar drie extra sterfgevallen aan pokken vinden onder de 32 mensen die op 24-jarige leeftijd nog vatbaar zijn. Samenvattend: vanaf 1.300 geboorten is het lot van 101 mensen om te sterven aan pokken. Dit is bijna precies de verwachte fractie $1/13$.

Bernoulli beschouwde toen de situatie waarin pokken bij de geboorte bij iedereen zouden worden ingeënt en geen sterfgevallen zouden veroorzaken. De pokken zouden uitgeroeid zijn en de vraag is de toename van de levensverwachting te schatten. Laten we, uitgaande van hetzelfde aantal geboorten P_0 , $P^*(x)$ het aantal mensen van x noemen wanneer de pokken zijn verdwenen.

¹Het feit dat p en q gelijk zijn is gewoon toeval.

Dan is

$$\frac{dP^*}{dx} = -m(x)P^*. \quad (4.6)$$

Bernoulli kon aantonen dat

$$P^*(x) = \frac{P(x)}{1 - p + pe^{-qx}}, \quad (4.7)$$

waarbij $P(x)$ zoals hierboven vermeld het cohort met leeftijd x in de aanwezigheid van pokken.

Door zoals eerder $m(x)$ tussen de vergelijkingen (4.6) en (4.3) te elimineren, verkrijgt Bernoulli na herschikking

$$\frac{1}{P^*} \frac{dP}{dx} - \frac{P}{P^{*2}} \frac{dP^*}{dx} = -pq \frac{S}{P} \frac{P}{P^*}.$$

Hij stelde $h(x) = P(x)/P^*(x)$. Met behulp van formule (4.4) vermenigvuldigde hij teller en noemer met e^{-qx} en verkreeg

$$\frac{1}{h} \frac{dh}{dx} = -pq \frac{e^{-qx}}{1 - p + pe^{-qx}},$$

wat gelijkwaardig is aan $\frac{d}{dx} \log h = \frac{d}{dx} \log(1 - p + pe^{-qx})$, waarbij \log hier staat voor de natuurlijke logaritme en niet voor de decimale logaritme. Maar $h(0) = 1$. Dus $h(x) = 1 - p + pe^{-qx}$.

Merk op dat de verhouding $P(x)/P^*(x)$ naar $1 - p$ neigt als de leeftijd x hoog genoeg is. De zesde kolom van tabel 4.1 toont $P^*(x)$. Een manier om $P(x)$ en $P^*(x)$ te vergelijken is het schatten van de levensverwachting bij de geboorte, waarvan de theoretische uitdrukking met pokken is

$$\frac{1}{P_0} \int_0^{+\infty} P(x) dx.$$

Een soortgelijke uitdrukking met $P^*(x)$ in plaats van $P(x)$ geldt ook zonder pokken. Bernoulli gebruikte de benaderende formule $[\frac{1}{2}P(0) + P(1) + P(2) + \dots]/P_0$, die gegeven wordt door de methode van de trapezoiden (figuur 4.2). Als men de tabel verderzet na 24 jaar tot 84 jaar (zie tabel 2.1), verkrijgt men uiteindelijk een levensverwachting E met pokken gelijk aan $[\frac{1}{2} \times 1.300 + 1.000 + \dots + 20]/1.300 \approx 26,57$ jaar, d.i. 26 jaar en 7 maanden. Zonder pokken had hij een levensverwachting van E^* gelijk aan $[\frac{1}{2} \times 1.300 + 1.015 + \dots + 23]/1.300 \approx 29,65$ jaar, d.i. 29 jaar en 8 maanden. Inenting bij de geboorte zou de levensverwachting met meer dan drie jaar doen toenemen.

We kunnen opmerken dat er een eenvoudiger en snellere methode is dan die van Bernoulli om deze formules te verkrijgen. Uitgaande van de differentiaalvergelijking (4.1) voor $S(x)$, zien we eerst dat

$$S(x) = P_0 e^{-qx} \exp\left(-\int_0^x m(y) dy\right).$$

Gebruik makend van deze uitdrukking in vergelijking (4.2) voor $R(x)$, vinden we dat

$$R(x) = P_0 (1-p) (1 - e^{-qx}) \exp\left(-\int_0^x m(y) dy\right).$$

Vergelijking (4.6) voor $P^*(x)$ toont aan dat

$$P^*(x) = P_0 \exp\left(-\int_0^x m(y) dy\right). \quad (4.8)$$

De formules (4.4) en (4.7) volgen onmiddellijk!

Natuurlijk is inenting met een minder virulente pokkenstam niet helemaal veilig. Als p' de kans is om vlak na de inenting aan de pokken te sterven ($p' < p$), dan zou de levensverwachting $(1 - p')E^*$ zijn als iedereen bij de geboorte zou worden ingeënt. Deze levensverwachting blijft hoger dan de natuurlijke levensverwachting E indien $p' < 1 - E/E^*$ of ongeveer 11 %. Gegevens over p' waren in die tijd moeilijk te verkrijgen. Maar Bernoulli schatte dat het risico p' kleiner was dan 1 %. Voor hem was er geen twijfel mogelijk: inenting moest door de staat worden bevorderd. Hij concludeerde:

„Ik zou alleen willen dat, in een zaak die zo nauw verband houdt met het welzijn van het menselijk ras, geen beslissing wordt genomen zonder alle kennis die een beetje analyse en berekening kunnen verschaffen.”

Bernoulli's werk werd in april 1760 gepresenteerd aan de Academie van Wetenschappen in Parijs. In november presenteerde d'Alembert een commentaar getiteld *Over de toepassing van de waarschijnlijkheidsleer op de inenting van pokken*. Het commentaar werd kort daarna gepubliceerd in het tweede deel van zijn *Opuscles mathématiques* met meer gedetailleerde berekeningen en samen met een ander werk getiteld *Wiskundige theorie van inenting*. D'Alembert bekritiseerde Bernoulli's aannames dat de kans op infectie en de kans op overlijden aan pokken onafhankelijk zijn van de leeftijd.

Hij stelde een andere oplossing voor die deze veronderstellingen niet vereist. Noem $v(x)$ de sterfte door pokken op leeftijd x , $m(x)$ de sterfte door andere oorzaken en $P(x)$ het aantal mensen dat nog in leven is. Dan is

$$\frac{dP}{dx} = -v(x)P - m(x)P. \quad (4.9)$$

Bij vergelijking met vergelijking (4.3) zien we dat in feite $v(x) = pqS(x)/P(x)$. Hier krijgen we

$$P^*(x) = P(x) \exp\left(\int_0^x v(y) dy\right), \quad (4.10)$$

waarbij $P^*(x)$ staat voor het aantal mensen dat nog leeft op leeftijd x wanneer de pokken zijn verdwenen.



Figuur 4.3:
D'Alembert (1717–1783)

Inderdaad kunnen we ofwel de functie $m(x)$ tussen vergelijkingen (4.6) en (4.9) substitueren ofwel formule (4.8) gebruiken voor $P^*(x)$ en opmerken dat de oplossing van (4.9) gegeven wordt door $P(x) = P_0 \times \exp(-\int_0^x [v(y) + m(y)] dy)$.

De formule van d'Alembert (4.10) is niet in tegenspraak met de formule van Bernoulli (4.7). Hij gebruikt alleen een ander soort informatie $v(x)$, die toen niet beschikbaar was omdat de overlijdensregisters wel de doodsoorzaak vermeldden maar niet de leeftijd van het slachtoffer. D'Alembert suggereerde dat men niet echt kon concluderen of inenting nuttig was voordat dit soort gegevens beschikbaar werd.

D'Alembert bekritiseerde ook de bruikbaarheid van de levensverwachting als beslissingscriterium, aangezien deze aan alle jaren hetzelfde gewicht toekent, ongeacht of ze in de nabije of de verre toekomst liggen. Hij merkte

op dat, vanuit het standpunt van het individu of van de Staat, niet alle jaren dezelfde „bruikbaarheid” hebben, waarbij jonge en oude leeftijden minder waardevol zijn dan gemiddelde leeftijden. Ondanks al deze kritiek verklaarde d’Alembert zich voorstander van inenting.

Door vertraging in de publicatie werd Bernoulli’s werk pas in 1766 gepubliceerd, terwijl d’Alembert er in slaagde zijn eigen werk zeer snel gepubliceerd te krijgen. Bernoulli uitte zijn verbittering in een brief aan Euler:

„Wat zegt u van de enorme platitudes van de grote d’Alembert over waarschijnlijkheden: omdat ik mij te vaak onrechtvaardig behandeld voel in zijn publicaties, heb ik al enige tijd geleden besloten niets meer te lezen wat uit zijn pen komt. Ik heb dit besluit genomen naar aanleiding van een manuscript over inenting dat ik acht jaar geleden naar de Academie in Parijs heb gestuurd en dat zeer gewaardeerd werd vanwege de nieuwigheid van de analyse. Het was, durf ik te zeggen, als het opnemen van een nieuwe tak in het corpus van de wiskunde. Het schijnt dat het succes van deze nieuwe analyse hem pijn in het hart heeft bezorgd. Hij heeft haar op duizend manieren bekritiseerd, allemaal even belachelijk, en nadat hij haar goed bekritiseerd heeft, doet hij alsof hij de eerste auteur is van een theorie die hij niet alleen heeft horen noemen. Hij wist echter dat mijn manuscript pas na zo’n zeven of acht jaar kon verschijnen. Hij kon er slechts kennis van hebben in zijn hoedanigheid van lid van de Academie. In dit opzicht had mijn manuscript heilig moeten blijven tot het openbaar werd gemaakt. *Dolus an virtus quis in hoste requirat?*”²

Ondanks de werken van Bernoulli en d’Alembert werd inenting in Frankrijk niet op grote schaal uitgevoerd. Koning Lodewijk XV stierf in 1774 aan de pokken. De artsen van het hof entten kort daarna de rest van de koninklijke familie in. Het probleem verloor aan belang toen Edward Jenner ontdekte dat het inenten van mensen met koepokken („vaccinatie”) beschermde tegen pokken en veilig was. Zijn werk, *Een onderzoek naar de oorzaken en gevolgen van het variolae-vaccin*, werd in 1798 gepubliceerd. Vaccinatie verspreidde zich snel over Europa. Niettemin zijn de methoden die werden ontwikkeld voor de berekening van de toename van de levensverwachting als één doodsoorzaak wordt weggenomen, nog steeds in gebruik.

In de daaropvolgende decennia kwamen er gegevens beschikbaar over de leeftijd waarop mensen aan pokken stierven. Het probleem werd opnieuw bekeken, vooral door

²List of dapperheid, wie zal daar tegenover een vijand om malen? Vergilius: *Aeneis*, Boek II.

- Johann Heinrich Lambert, een wiskundige van de Berlijnse Academie, in 1772;
- Emmanuel-Étienne Duvillard, toen belast met bevolkingsstatistiek op het ministerie van Binnenlandse Zaken in Parijs, in zijn *Analyse en tabellen van de invloed van pokken op de sterfte op elke leeftijd* (1806);
- Pierre-Simon Laplace in zijn *Analytische theorie der waarschijnlijkheden* (1812).

Duvillard en Laplace toonden bijvoorbeeld aan hoe de formule (4.7) kan worden gewijzigd wanneer de parameters p en q van de leeftijd afhangen:

$$P^*(x) = \frac{P(x)}{1 - \int_0^x p(y) q(y) e^{-\int_0^y q(z) dz} dy}.$$

Hierbij is $p(x)$ de kans om aan pokken te sterven als men op leeftijd x besmet is en $q(x)$ de kans om op leeftijd x met pokken besmet te zijn.

Na dit werk over pokken heeft Daniel Bernoulli zich niet meer beziggehouden met enig ander probleem in de populatiedynamica. Hij stierf in Basel in 1782. D'Alembert stierf een jaar later in Parijs.

Verdere lectuur

1. D'Alembert, J.: Sur l'application du calcul des probabilités à l'inoculation de la petite vérole. In: *Opuscles mathématiques*, II, 26–95 (1761). gallica.bnf.fr
2. Bernoulli, D.: Réflexions sur les avantages de l'inoculation. *Mercure de France*, 173–190 (juin 1760). retronews.fr
3. Bernoulli, D.: Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. *Hist. Acad. R. Sci. Paris*, 1–45 (1760/1766). gallica.bnf.fr
4. Dietz, K., Heesterbeek, J.A.P.: Daniel Bernoulli's epidemiological model revisited. *Math. Biosci.* 180, 1–21 (2002)
5. Duvillard, E.E.: *Analyse et tableaux de l'influence de la petite vérole sur la mortalité à chaque âge*. Imprimerie Impériale, Paris (1806). archive.org
6. Lambert, J.H.: *Contributions mathématiques à l'étude de la mortalité et de la nuptialité* (1765 et 1772). INED, Paris (2006).
7. Laplace, P.S.: *Théorie analytique des probabilités* (1812). gallica.bnf.fr
8. Straub, H.: Bernoulli, Daniel. In Gillespie, C.C. (ed.) *Dictionary of Scientific Biography*, vol. 2, 36–46. Scribner, New York (1970)
9. Tent, M.B.W.: *Leonhard Euler and the Bernoullis*. A K Peters, Natick (2009)
10. Voltaire: *Lettres philosophiques*. Lucas, Amsterdam (1734). gallica.bnf.fr

Hoofdstuk 5

Malthus en de belemmeringen voor geometrische groei (1798)

In 1798 publiceerde Malthus *Een essay over het principe van de bevolkingsgroei*, waarin hij betoogde dat het aanbod van voedsel de natuurlijke neiging van de menselijke bevolking om exponentieel te groeien niet voor langere tijd kon volgen. Als de bevolking relatief constant bleef, kwam dat omdat een groot deel van de mensheid te kampen had met voedselschaarste. Malthus zag het „beginsel van bevolking” als een argument tegen de geschriften van Godwin en Condorcet, die de nadruk legden op vooruitgang in menselijke samenlevingen. Malthus’ essay beïnvloedde de evolutietheorie van Darwin en Wallace en werd bekritiseerd door Marx, maar werd in praktijk gebracht met het Chinese één-kind beleid.

Thomas Robert Malthus werd in 1766 in de buurt van Londen geboren als zesde van zeven kinderen. Zijn vader, een vriend en bewonderaar van Jean-Jacques Rousseau, was zijn eerste leraar. In 1784 begon de jonge Malthus wiskunde te studeren aan de Universiteit van Cambridge. Hij behaalde zijn diploma in 1791, werd fellow van het *Jesus College* in 1793 en Anglicaans priester in 1797.



Figuur 5.1:
Malthus (1766–1834)

In 1798 publiceerde Malthus anoniem een boek getiteld *Een essay over*

het principe van de bevolkingsgroei, zoals het de toekomstige verbetering van de samenleving beïnvloedt, met opmerkingen over de speculaties van de heer Godwin, de heer Condorcet en andere schrijvers. Het kwam als een kritiek op Godwin's *Onderzoek naar politieke rechtvaardigheid* (1793) en Condorcet's *Schets van een historisch tafereel der vorderingen van 's menschen geest* (1794). Ondanks de verschrikkingen die de Franse Revolutie aanrichtte in naam van de vooruitgang, beweerden de twee auteurs dat de vooruitgang van de maatschappij onvermijdelijk was. Malthus deelde niet hetzelfde optimisme. Hij betoogde ook dat de Engelse *Poor Laws*, die arme gezinnen met veel kinderen hielpen, de groei van de bevolking bevorderden zonder een vergelijkbare groei in de productie van voedsel aan te moedigen. Het leek hem dat deze wetten de armen niet echt ontlastten, integendeel. Meer in het algemeen, omdat de bevolking altijd sneller groeide dan de voedselproductie, leek een deel van de maatschappij veroordeeld tot ellende, honger of epidemieën: dit zijn de plagen die de bevolkingsgroei vertragen en die, naar Malthus' mening, de voornaamste obstakels zijn voor de vooruitgang van de maatschappij. Alle theorieën die vooruitgang beloven zouden slechts utopisch zijn. Deze ideeën brachten Malthus ertoe zijn boek te publiceren in 1798. Hier is hoe hij zijn thesis samenvatte:

[...] „de kracht van de bevolking is oneindig veel groter dan de kracht van de aarde om in het levensonderhoud van de mens te voorzien. Bevolking, wanneer ongecontroleerd, neemt toe in een geometrische verhouding. De bestaansmiddelen nemen slechts toe in een rekenkundige verhouding. Een kleine kennismaking met getallen zal de onmetelijkheid van de eerste macht in vergelijking met de tweede aantonen. Door die wet van onze natuur die voedsel noodzakelijk maakt voor het leven van de mens, moeten de effecten van deze twee ongelijke krachten gelijk gehouden worden. Dit impliceert een sterke en voortdurend werkende controle op de bevolking door de problemen met de bestaansmiddelen. Dit probleem moet ergens optreden en moet noodzakelijkerwijs door een groot deel van de mensheid ernstig worden gevoeld.”

Malthus' boek was zeer succesvol. Het bevatte weinig gegevens. Malthus merkte bijvoorbeeld op dat de bevolking van de VS gedurende de achttiende eeuw elke vijftiende jaar was verdubbeld. Hij probeerde niet echt zijn stellingen te vertalen in wiskundige modellen maar effende de weg voor later werk van Adolphe Quetelet en Pierre-François Verhulst, die het onderwerp zullen zijn van het volgende hoofdstuk.

Na de publicatie van zijn boek reisde Malthus met vrienden eerst naar Duitsland, Scandinavië en Rusland, en vervolgens naar Frankrijk en Zwitserland. Hij gebruikte de tijdens zijn reizen verzamelde informatie om onder zijn naam een sterk uitgebreide tweede editie in 1803 te publiceren, met een andere ondertitel: *Een essay over het principe van de bevolkingsgroei, of een overzicht van de vroegere en huidige gevolgen voor het menselijk geluk, met een onderzoek naar onze vooruitzichten voor de toekomstige opheffing of beperking van de kwalen die erdoor worden veroorzaakt*. Deze nieuwe uitgave behandelde in detail de obstakels voor bevolkingsgroei in verschillende landen: uitgesteld huwelijk, abortus, kindermoord, hongersnood, oorlog, epidemieën, economische factoren. Voor Malthus was het uitstellen van het huwelijk de beste optie om de bevolking te stabiliseren. Er volgden nog vier edities van het boek in 1806, 1807, 1817 en 1826. In 1805 werd Malthus professor in geschiedenis en politieke economie in een nieuwe school die door de West Indies Company was opgezet voor haar werknemers. Hij publiceerde ook *Een onderzoek naar de aard en de vooruitgang van de huur* (1815) en *Beginselen van de politieke economie* (1820). In 1819 werd Malthus gekozen tot lid van de *Royal Society*. In 1834 was hij een van de oprichters van het Statistisch Genootschap. Hij stierf in de buurt van Bath in datzelfde jaar.

Het werk van Malthus had een sterke invloed op de ontwikkeling van de evolutietheorie. Charles Darwin, terug van zijn reis aan boord van de *Beagle*, las Malthus' boek over bevolking in 1838. Dit is wat hij schreef in de inleiding van zijn beroemde boek *Over het ontstaan van soorten door natuurlijke selectie*, gepubliceerd in 1859:

„In het volgende hoofdstuk zullen wij de Strijd om het Bestaan tussen alle organische wezens op de wereld behandelen, die onvermijdelijk het gevolg is van hun grote geometrische kracht van vermeerdering. Dit is de leer van Malthus, toegepast op het dierenrijk en het plantenrijk.”

Alfred Russel Wallace, die tegelijk met Darwin de evolutietheorie ontwikkelde, zei ook dat zijn ideeën ontstonden na het lezen van Malthus' boek.

In tegenstelling hiermee is hier het standpunt van Karl Marx over het succes van Malthus' boek, zoals te lezen is in een voetnoot van zijn *Kapitaal*:

„Misschien denkt de lezer hierbij aan Malthus, wiens *Essay on Population* in 1798 verscheen. Maar ik herinner er aan dat dit geschrift in zijn eerste vorm niets anders is dan een onbeholpen, oppervlakkig en domineeachtig plagiaat uit de werken van De Foe, Sir James Steuart, Townsend, Franklin, Wallace, enzovoort, en geen enkele zelfbedachte zin bevat. Het was alleen maar aan

partijbelangen te danken dat dit pamflet zulk een groot opzien baarde. De Franse Revolutie had in het Britse Rijk hartstochtelijke verdedigers gevonden; het ‚bevolkingsbeginsel’, dat in de achttiende eeuw langzamerhand was uitgewerkt en vervolgens in een grote sociale crisis met veel lawaai werd verkondigd als het onfeilbare tegengif voor de leer van Condorcet en anderen, werd door de Engelse oligarchie jubelend begroet als de grote verdeler van iedere begeerte naar een verdere ontwikkeling van de mens. Malthus, ten zeerste verbaasd over zijn succes, begon nu oppervlakkig verzameld materiaal in het oude schema te stoppen en er nieuwe dingen - die echter niet door Malthus waren ontdekt, maar slechts door hem waren toegeëigend - aan toe te voegen.”

De stellingen van Malthus waren zeker niet geheel nieuw. Bijvoorbeeld, het idee dat de bevolking de neiging heeft om geometrisch te groeien wordt vaak aan hem toegeschreven, hoewel we in Hoofdstuk 3 zagen dat dit idee al een halve eeuw eerder bekend was bij Euler¹. Malthus gaf er echter bekendheid aan door het op polemische wijze te verbinden met reële wetgevingsproblemen. Ironisch genoeg was het in communistisch China dat Malthus’ voorstel om geboorten te beperken zijn meest opvallende toepassing zou vinden (zie hoofdstuk 25).

Verdere lectuur

1. Condorcet: *Schets van een historisch tafereel der vorderingen van ’s menschen geest*. Haarlem (1802). condorcet.nl
2. Darwin, C. (Vertaler : Winkler, T. C.) *Het ontstaan der soorten van dieren en planten door middel van de natuurkeus*. A. C. Kruseman, Haarlem (1860). archive.org
3. Godwin, W.: *An Enquiry Concerning Political Justice* (1793). archive.org
4. Malthus, T.R.: *An Essay on the Principle of Population* (1798). econlib.org
5. Marx, K. (Vertaler : Lipschits, I.) *Het Kapitaal. Een kritische beschouwing van de economische politiek*, C. de Boer jr., Paul Brand, Hilversum - Antwerpen (1967) marxists.org
6. Simpkins, D.M.: Malthus, Thomas Robert. In: Gillespie, C.C. (ed.) *Dictionary of Scientific Biography*, vol. 9, 67–71. Scribner, New York (1974)

¹R. A. Fisher (zie hoofdstukken 14 en 20) zou de „Malthusiaanse parameter” de groeisnelheid van bevolkingen noemen. Malthus noemde in zijn eigen boek de verhandeling van Süßmilch.

Hoofdstuk 6

Verhulst en de logistische vergelijking (1838)

In 1838 introduceerde de Belgische wiskundige Verhulst de logistische vergelijking, die een soort veralgemening is van de vergelijking voor exponentiële groei, maar met een maximumwaarde voor de bevolking. Hij gebruikte gegevens van verschillende landen, met name België, om de onbekende parameters te schatten. Het werk van Verhulst werd pas in de jaren 1920 herontdekt.

Pierre-François Verhulst werd in 1804 in Brussel geboren. Hij promoveerde in 1825 in de wiskunde aan de Universiteit van Gent. Hij was ook geïnteresseerd in politiek. Toen hij in Italië was om zijn tuberculose te beteu- gelen, pleitte hij zonder succes voor een grondwet voor de Pauselijke Staten. Na de revolutie van 1830 en de onafhankelijkheid van België publiceerde hij een historisch essay over een achttiende-eeuwse patriot. In 1835 werd hij professor in de wiskunde aan de nieuw opgerichte Vrije Universiteit in Brussel.



Figuur 6.1:
Verhulst (1804–1849)

In datzelfde jaar 1835 publiceerde zijn landgenoot Adolphe Quetelet, een statisticus en directeur van het observatorium in Brussel, *Een verhandeling over de mens en de ontwikkeling van zijn vermogens*. Quetelet suggereerde dat bevolkingen niet meetkundig konden groeien over een lange periode omdat de door Malthus genoemde obstakels een soort „weerstand” vormden, die volgens hem (naar analogie van de mechanica) evenredig was met het kwadraat van de snelheid van bevolkingsgroei. Deze analogie had geen echte

basis, maar zij inspireerde Verhulst.

Verhulst publiceerde in 1838 een *Nota over de wet van de bevolkingsgroei*. Hier zijn enkele uittreksels:

„Wij weten dat de beroemde Malthus het principe heeft aange-
toond dat de menselijke bevolking de neiging heeft om in een ge-
ometrische progressie te groeien om na een bepaalde tijd te ver-
dubbelen, bijvoorbeeld om de vijftwintig jaar. Deze stelling is
onbetwistbaar als men abstraheert van de toenemende moeilijk-
heid om voedsel te vinden [...]

De virtuele toename van de bevolking wordt dus beperkt door de
omvang en de vruchtbaarheid van het land. Als gevolg daarvan
komt de bevolking steeds dichter bij een evenwichtstoestand.”

Verhulst beseft waarschijnlijk dat Quetelet's mechanische analogie niet re-
delijk was en stelde in plaats daarvan de volgende (nog steeds enigszins arbi-
traire) differentiaalvergelijking voor de populatie $P(t)$ op tijdstip t voor:

$$\frac{dP}{dt} = rP \left(1 - \frac{P}{K}\right). \quad (6.1)$$

Wanneer de populatie $P(t)$ klein is in vergelijking met de parameter K , krij-
gen we bij benadering de vergelijking $\frac{dP}{dt} \approx rP$, waarvan de oplossing $P(t) \approx$
 $P(0)e^{rt}$ is, d.w.z. exponentiële groei¹. De groeisnelheid neemt af naarmate
 $P(t)$ dichterbij K komt. Ze zou zelfs negatief worden als $P(t)$ K zou over-
schrijden. Om de exacte uitdrukking van de oplossing van vergelijking (6.1)
te krijgen, kunnen we te werk gaan zoals Daniel Bernoulli voor vergelijking
(4.5).

Door vergelijking (6.1) te delen door P^2 en $p = 1/P$ te stellen, krijgen
we $\frac{dp}{dt} = -rp + r/K$. Met $q = p - 1/K$ krijgen we $\frac{dq}{dt} = -rq$ en $q(t) =$
 $q(0)e^{-rt} = (1/P(0) - 1/K)e^{-rt}$. Dus kunnen we $p(t)$ en $P(t)$ afleiden.

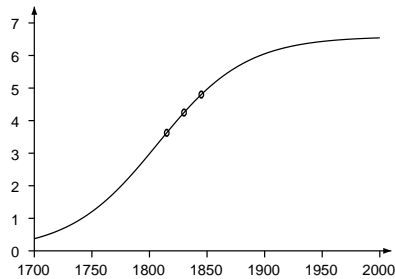
Uiteindelijk krijgen we na herschikking

$$P(t) = \frac{P(0)e^{rt}}{1 + P(0)(e^{rt} - 1)/K}. \quad (6.2)$$

De totale populatie neemt geleidelijk toe van $P(0)$ op het tijdstip $t = 0$ tot de
limiet K , die pas wordt bereikt als $t \rightarrow +\infty$ (figuur 6.2).

¹Gewoonlijk spreekt men van geometrische groei in discrete-tijdmodellen en van exponentiële groei in continue-tijdmodellen, maar het is in wezen hetzelfde.

Figuur 6.2: De bevolking van België (in miljoenen) en de logistische curve. De datapunten komen overeen met de jaren 1815, 1830 en 1845. De parameterwaarden zijn die van het artikel uit 1845.



Zonder de waarden te geven die hij gebruikte voor de onbekende parameters r en K , vergeleek Verhulst zijn resultaat met gegevens over de bevolking van Frankrijk tussen 1817 en 1831, van België tussen 1815 en 1833, van het graafschap Essex in Engeland tussen 1811 en 1831, en van Rusland tussen 1796 en 1827. Het resultaat bleek vrij goed te zijn.

In 1840 werd Verhulst professor aan de Koninklijke Militaire School te Brussel. Het jaar daarop publiceerde hij een *Elementaire Verhandeling over Elliptische Functies* en werd hij verkozen tot lid van de Koninklijke Academie van België. In 1845 zette hij zijn studie van de bevolking voort met een artikel getiteld *Mathematische onderzoekingen over de wet van de bevolkingsgroei*. Hij greep eerst terug op de opmerking van Malthus volgens wie de bevolking van de Verenigde Staten elke 25 jaar verdubbeld was (tabel 6.1).

Tabel 6.1: Officiële volkstellingen van de VS.

jaar	bevolking	jaar	bevolking
1790	3.929.827	1820	9.638.131
1800	5.305.925	1830	12.866.020
1810	7.239.814	1840	17.062.566

Als we de verhouding berekenen tussen de bevolking in het jaar $n + 10$ en die in het jaar n , vinden we respectievelijk 1,350, 1,364, 1,331, 1,335 en 1,326, wat vrij constant is. De bevolking is dus gemiddeld elke 10 jaar met 1,34 vermenigvuldigd en elke 25 jaar met $1,34^{25/10} \approx 2,08$. De bevolking was dus elke 25 jaar blijven verdubbelen sinds het essay van Malthus, bijna een halve eeuw eerder. Verhulst voegde er echter aan toe:

„Wij zullen niet aandringen op de hypothese van de geometrische progressie, aangezien deze slechts in zeer bijzondere om-

standigheden kan gelden; bijvoorbeeld wanneer een vruchtbaar gebied van bijna onbeperkte omvang toevallig wordt bewoond door mensen met een geavanceerde beschaving, zoals het geval was voor de eerste Amerikaanse koloniën.”

In zijn artikel kwam Verhulst ook terug op vergelijking (6.1), die hij „logistisch” noemde. Hij merkte op dat de kromme $P(t)$ toeneemt met een positieve kromming (ze is convex) zolang $P(t) < K/2$ en dan blijft toenemen naar K toe maar met een negatieve kromming (ze is concaaf) zodra $P(t) > K/2$. De kromme heeft dus de vorm van een vervormde letter S (figuur 6.2).

Inderdaad, $\frac{d^2P}{dt^2} = r(1 - 2P/K) \frac{dP}{dt}$. Dus $\frac{d^2P}{dt^2} > 0$ als $P < K/2$ en $\frac{d^2P}{dt^2} < 0$ als $P > K/2$.

Verhulst legde ook uit hoe de parameters r en K kunnen worden geschat op basis van de populatie $P(t)$ in drie verschillende maar even ver uit elkaar liggende jaren. Als P_0 de populatie is op het tijdstip $t = 0$, P_1 die op het tijdstip $t = T$ en P_2 die op het tijdstip $t = 2T$, dan leert een omslachtige berekening uitgaande van vergelijking (6.2) dat

$$K = P_1 \frac{P_0 P_1 + P_1 P_2 - 2 P_0 P_2}{P_1^2 - P_0 P_2}, \quad r = \frac{1}{T} \log \left[\frac{1/P_0 - 1/K}{1/P_1 - 1/K} \right].$$

Op basis van de schattingen voor de bevolking van België in de jaren 1815, 1830 en 1845 (respectievelijk 3,627, 4,247 en 4,801 miljoen), berekende hij $K = 6,584$ miljoen en $r = 2,6\%$ per jaar. Met de vergelijking (6.2) kon hij dan voorspellen dat de bevolking van België 4,998 miljoen zou bedragen bij het begin van het jaar 1851 en 6,064 miljoen bij het begin van het jaar 1900 (figuur 6.2). Verhulst deed een soortgelijke studie voor Frankrijk. Hij berekende $K = 39,685$ miljoen en $r = 3,2\%$ per jaar. Aangezien de bevolking van België en Frankrijk deze waarden van K intussen ruimschoots overschreden heeft, zien we dat de logistische vergelijking slechts een realistisch model kan zijn voor periodes van enkele tientallen jaren, zoals in het artikel van Verhulst van 1838, maar niet voor langere periodes.

In 1847 verscheen een *Tweede onderzoek naar de wet van de bevolkingsgroei* waarin Verhulst de logistische vergelijking opgaf en in plaats daarvan koos voor een differentiaalvergelijking die kan worden geschreven in de vorm $\frac{dP}{dt} = r(1 - P/K)$. Hij dacht dat deze vergelijking zou opgaan wanneer de populatie $P(t)$ boven een bepaalde drempel ligt. De oplossing is $P(t) = K + (P(0) - K)e^{-rt/K}$. Aan de hand van dezelfde demografische gegevens voor België schatte Verhulst opnieuw de parameters r en K . Ditmaal

vond hij $K = 9,4$ miljoen voor de maximale bevolking. We zien hoezeer het resultaat kan afhangen van de keuze van het model!

Verhulst werd voorzitter van de Koninklijke Academie van België in 1848, maar overleed het jaar daarop in Brussel, waarschijnlijk aan tuberculose. Ondanks Verhulst's aarzeling tussen modelvergelijkingen, werd de logistische vergelijking enkele decennia later door verschillende mensen onafhankelijk opnieuw geïntroduceerd. Robertson gebruikte haar in 1908 om de individuele groei van dieren, planten, mensen en lichaamsorganen te modelleren. McKendrick en Kesava Pai gebruikten haar in 1911 voor de groei van populaties van micro-organismen. Pearl en Reed gebruikten het in 1920 voor de groei van de bevolking van de VS, die was begonnen af te nemen. In 1922 merkte Pearl eindelijk het werk van Verhulst op. Vanaf dat moment inspireerde de logistische vergelijking tot vele werken (zie de hoofdstukken 13, 20 en 24). De maximale populatie K werd uiteindelijk bekend als de „draagkracht”.

Verdere lectuur

1. Lloyd, P.J.: American, German and British antecedents to Pearl and Reed's logistic curve. *Pop. Stud.* 21, 99–108 (1967)
2. McKendrick, A.G., Kesava Pai, M.: The rate of multiplication of micro-organisms: A mathematical study. *Proc. R. Soc. Edinb.* 31, 649–655 (1911)
3. Pearl, R.: *The Biology of Death*. Lippincott, Philadelphia (1922). archive.org
4. Pearl, R., Reed, L.J.: On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc. Natl. Acad. Sci.* 6, 275–288 (1920). pnas.org
5. Quetelet, A.: *Sur l'homme et le développement de ses facultés*. Bachelier, Paris (1835). gallica.bnf.fr
6. Quetelet, A.: Pierre-François Verhulst. *Annu. Acad. R. Sci. Lett. B.-Arts Belg.* 16, 97–124 (1850). archive.org
7. Quetelet, A.: *Sciences mathématiques et physiques au commencement du XIXe siècle*. Mucquardt, Bruxelles (1867). gallica.bnf.fr
8. Robertson, T.B.: On the normal rate of growth of an individual and its biochemical significance. *Arch. Entwicklungsmechanik Org.* 25, 581–614 (1908)
9. Verhulst, P.-F.: Notice sur la loi que la population poursuit dans son accroissement. *Corresp. Math. Phys.* 10, 113–121 (1838). archive.org
10. Verhulst, P.-F.: Recherches mathématiques sur la loi d'accroissement de la population. *Nouv. Mém. Acad. R. Sci. B.-lett. Brux.* 18, 1–45 (1845). uni-goettingen.de
11. Verhulst, P.-F.: Deuxième mémoire sur la loi d'accroissement de la population. *Mém. Acad. R. Sci. Lett. B.-Arts Belg.* 20 (1847). archive.org

Hoofdstuk 7

Bienaymé, Cournot en het uitsterven van familienamen (1845–1847)

De Franse statisticus Bienaymé begreep in 1845 hoe hij de kans kon berekenen dat een familienaam zou uitsterven als elke man een aantal zonen had volgens een gegeven kansverdeling. Als het gemiddelde aantal zonen kleiner is dan of gelijk aan één, zal de familienaam uitsterven. Als het gemiddelde groter is dan één, is de uitsterfkans strikt kleiner dan één. Het bewijs van zijn resultaat werd twee jaar later gepubliceerd in een boek geschreven door zijn vriend Cournot. Deze werken werden pas onlangs herontdekt.

Irenée Jules Bienaymé werd in 1796 in Parijs geboren. Hij studeerde aan de *École Polytechnique* en maakte carrière op het Ministerie van Financiën, waar hij de hoge positie van inspecteur-generaal bereikte. Geïnspireerd door het boek van Laplace *Analytische theorie der waarschijnlijkheden* vond Bienaymé ook nog tijd om artikelen te publiceren over talrijke toepassingen van kansrekening, zoals demografische en medische statistiek (kindersterfte, aantal geboorten, levensverwachting), de kans op fouten in de rechtspraak, verzekeringstheorie en representativiteit van stelsystemen.



Figuur 7.1:
Bienaymé (1796–1878)

In 1845 schreef Bienaymé een kort stukje *Over de wet van vermenigvuldiging en de duur van families*, dat werd gepubliceerd in het bulletin van de *Société Philomatique* in Parijs. Een aantal auteurs had reeds over dit onderwerp geschreven. In de tweede editie van *Een essay over het principe van de bevolkingsgroei* (1803) nam Malthus een hoofdstuk op over de bevolking van Zwitserland en merkte op dat

„in de stad Bern had de soevereine raad van 1583 tot 1654 487 families tot de burgerij toegelaten, waarvan er in twee eeuwen tijd 379 waren uitgestorven en er in 1783 nog slechts 108 over waren.”

In 1842 beweerde Thomas Doubleday meer in het algemeen dat adellijke of burgerlijke families uit de hogere klasse een grotere neiging hadden te verdwijnen dan families uit de lagere klasse. Soortgelijke ideeën werden in Frankrijk naar voren gebracht door Émile Littré in 1844 in een inleidende tekst bij de positivistische filosofie van Auguste Comte en door Benoiston de Châteauneuf – een vriend van Bienaymé – die in 1845 een essay publiceerde *Over de duur van adellijke families in Frankrijk*.

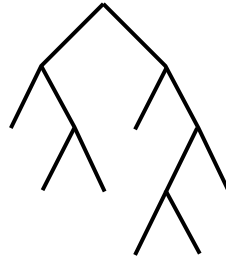
Het was in deze context dat Bienaymé probeerde te verklaren hoe het kon dat de bevolking van een land vaak geometrisch groeit terwijl een groot aantal families uitsterft. Om dit probleem aan te pakken ging hij uit van het vereenvoudigde geval waarin alle mannen dezelfde kans zouden hebben om 0, 1, 2, 3,... zonen te hebben die de volwassen leeftijd bereiken. In het bijzonder vroeg hij zich af hoe groot de kans voor een man was om na n generaties nog nakomelingen te hebben die zijn naam dragen. Als het gemiddelde aantal zonen kleiner is dan één, is het duidelijk dat deze kans naar nul moet gaan als n oneindig groot wordt. Bienaymé merkte op dat dezelfde conclusie¹ zou gelden indien het gemiddeld aantal zonen precies één was, bijvoorbeeld als er een kans $1/2$ is op geen zoon en een kans $1/2$ op twee zonen (fig. 7.2). Maar in dat geval gaat de kans op nakomelingen in generatie n langzamer naar nul: in het voorbeeld zou die nog steeds 5% bedragen na 35 generaties, d.w.z. na elf of twaalf eeuwen als er drie generaties per eeuw zijn². Tenslotte merkte Bienaymé op dat als het gemiddelde aantal zonen groter is dan één, het uitsterven van de familielijn niet zeker is: de kans dat dit gebeurt kan worden berekend door een bepaalde algebraïsche vergelijking op te lossen.

Het artikel van Bienaymé bevatte geen nadere uitleg. In 1847 nam zijn vriend Antoine-Augustin Cournot, een wiskundige en econoom, enkele details op in een boek getiteld *Over de oorsprong en de grenzen van de over-*

¹Behalve als iedere man precies één zoon heeft.

²Zoals we hierna zullen zien, is deze kans gelijk aan $1 - x_{35}$ met $x_{n+1} = \frac{1}{2} + \frac{1}{2}x_n^2$ en $x_0 = 0$.

Figuur 7.2: Voorbeeld van een stamboom. De stamvader staat bovenaan in de boom. In elke generatie hebben mannen een kans van $1/2$ om geen zoon te hebben en een kans van $1/2$ om twee zonen te hebben.



eenstemming tussen algebra en meetkunde. Hij presenteerde het probleem in de context van een kansspel, maar erkende dat het identiek was aan het onderzoek van Bienaymé naar het uitsterven van familienamen. Als we de interpretatie in termen van familienamen aanhouden, beschouwde Cournot eerst het speciale geval waarin mannen ten hoogste twee zonen hebben, waarbij p_0 , p_1 en p_2 respectievelijk de kansen op het hebben van 0, 1 of 2 zonen zijn. Uiteraard is $p_0 + p_1 + p_2 = 1$. Uitgaande van één voorouder is de kans op uitsterven na slechts één generatie, noem het x_1 , natuurlijk gelijk aan p_0 . De kans op uitsterven binnen twee generaties is $x_2 = p_0 + p_1 x_1 + p_2 x_1^2$: ofwel was de familie reeds in de eerste generatie uitgestorven (kans p_0), ofwel was er slechts één zoon in de eerste generatie die geen mannelijke nakomelingen had (kans $p_1 x_1$), ofwel waren er twee zonen in de eerste generatie en elk van hen had geen mannelijke nakomelingen (kans $p_2 x_1^2$). Meer algemeen is de kans op uitsterven binnen n generaties $x_n = p_0 + p_1 x_{n-1} + p_2 (x_{n-1})^2$. Indien er bijvoorbeeld twee zonen zijn in de eerste generatie (kans p_2), zal de familie $n - 1$ generaties later (dus in generatie n) uitsterven met een kans gelijk aan $(x_{n-1})^2$. Cournot merkte op dat x_n een stijgende reeks is met $x_n \leq 1$ voor alle n . Dus x_n heeft een limiet $x_\infty \leq 1$, die een oplossing is van de vergelijking $x = p_0 + p_1 x + p_2 x^2$. Met $p_1 = 1 - p_0 - p_2$ is deze vergelijking equivalent aan $0 = p_2(x - 1)(x - p_0/p_2)$. Er zijn dus twee wortels: $x = 1$ en $x = p_0/p_2$. Afhankelijk van het gemiddeld aantal zonen $p_1 + 2p_2$, dat ook gelijk is aan $1 - p_0 + p_2$ en dat we \mathcal{R}_0 zullen noemen, kunnen we drie gevallen onderscheiden. Als $\mathcal{R}_0 < 1$, dan is $p_0/p_2 > 1$, dus $x = 1$ is in dit geval de enig mogelijke waarde voor de limiet x_∞ en de familienaam sterft met zekerheid uit. Als $\mathcal{R}_0 = 1$, zijn beide wortels gelijk aan 1 en is de conclusie hetzelfde. Als $\mathcal{R}_0 > 1$, dan beredeneerde Cournot dat x_∞ gelijk moest zijn aan de tweede wortel p_0/p_2 , aangezien de uitsterfkans natuurlijk 0 moet zijn in het speciale geval waarin $p_0 = 0$.

Cournot vermeldde kort het meer algemene geval waarin mannen ten hoogste m zonen kunnen hebben, met kansen p_0, p_1, \dots, p_m . De conclusie

hangt op dezelfde manier af van de waarde van $\mathcal{R}_0 = p_1 + 2p_2 + \dots + mp_m$, het gemiddeld aantal zonen, ten opzichte van 1. De vergelijking voor x_∞ , die in dit geval $x = p_0 + p_1 x + \dots + p_m x^m$ is, heeft altijd de wortel $x = 1$. Zij heeft slechts één andere positieve wortel, die de uitsterfkans x_∞ geeft wanneer $\mathcal{R}_0 > 1$.

Helaas zijn het artikel van Bienaymé en de paar bladzijden in het boek van Cournot destijds totaal onopgemerkt gebleven. Het artikel werd pas in de jaren zeventig opgemerkt en de boekpagina's nog eens twintig jaar later! Intussen waren het probleem en de oplossing ervan door anderen herontdekt en had het onderwerp een aanzienlijke ontwikkeling doorgemaakt. We zullen daarop terugkomen in de hoofdstukken 9, 17 en 18.

Bienaymé moest zijn baan bij het Ministerie van Financiën opgeven na de revolutie van 1848. Ook de leerstoel voor waarschijnlijkheidsrekening aan de universiteit van Parijs, waarvoor hij zeker de beste kandidaat was, werd aan iemand anders gegeven. Toch kon Bienaymé na 1850 weer aan de slag bij het Ministerie van Financiën, maar hij nam in 1852 ontslag. Later dat jaar werd hij benoemd tot lid van de Academie van Wetenschappen, als specialist op het gebied van de statistiek. In 1853 bewees hij wat sommige moderne leerboeken de „Bienaymé-Tchebychev” ongelijkheid noemen. In 1875 werd hij voorzitter van de pas opgerichte *Société Mathématique de France*. Hij overleed in Parijs in 1878.

Verdere lectuur

1. Bienaymé, I.J.: De la loi de multiplication et de la durée des familles. *Extr. p. v. séances - Soc. Philomat. Paris*, 37–39 (1845) biodiversitylibrary.org
2. Bru, B.: À la recherche de la démonstration perdue de Bienaymé. *Math. Sci. Hum.* 114, 5–17 (1991). archive.numdam.org
3. Bru, B., Jongmans, F., Seneta, E.: I.J. Bienaymé: Family information and proof of the criticality theorem. *Int. Stat. Rev.* 60, 177–183 (1992)
4. Cournot, A.-A.: *De l'origine et des limites de la correspondance entre l'algèbre et la géométrie*. Hachette, Paris (1847). archive.org
5. Doubleday, T.: *The True Law of Population* (1842). archive.org
6. Heyde, C.C., Seneta, E.: *I.J. Bienaymé*. Springer (1977)
7. Kendall, D.G.: The genealogy of genealogy: branching processes before (and after) 1873. *Bull. Lond. Math. Soc.* 7, 225–253 (1975)
8. Littré, É.: *Conservation, révolution et positivisme* (1852). gallica.bnf.fr
9. Malthus, T.R.: *An Essay on the Principle of Population* (1803). archive.org
10. Martin, T.: Antoine Augustin Cournot. In: Heyde, C.C., Seneta, E. (eds.) *Statisticians of the Centuries*, 152–156. Springer (2001)
11. Seneta, E.: Irenée-Jules Bienaymé. In: *ibid.*, 132–136.

Hoofdstuk 8

Mendel en erfelijkheid (1865)

In 1865 publiceerde Mendel de resultaten van zijn baanbrekende experimenten over de hybridisatie van erwten. Zijn analyse maakte gebruik van elementaire aspecten van de kanstheorie. Hij bedacht ook een dynamisch model voor een populatie van zelfbestuivende planten. Zijn werk, dat pas in 1900 werd herontdekt, is een mijlpaal in de geschiedenis van de genetica.

Johann Mendel werd in 1822 geboren in Moravië, destijds een deel van het Oostenrijkse keizerrijk en vandaag een deel van Tsjechië. Zijn vader was een landbouwer. Met zijn goede resultaten op de middelbare school en zijn zwakke gezondheid, gaf Mendel er de voorkeur aan verder te studeren in plaats van op de familieboerderij te werken. Maar hij kon het zich niet veroorloven om naar de universiteit te gaan. Daarom trad hij in 1843 toe tot de abdij van Sint Thomas in Brünn (nu Brno), waar hij de naam Gregor aannam. Hij studeerde theologie, maar volgde ook enkele cursussen over landbouw. In 1847 werd hij tot priester gewijd. Hij gaf een paar jaar les aan een middelbare school, maar slaagde niet voor het examen om gewoon hoogleraar te worden. Tussen 1851 en 1853 kon hij dankzij de steun van zijn hiërarchie toch zijn studie voortzetten aan de universiteit van Wenen, waar hij vakken volgde in natuurkunde, wiskunde en natuurwetenschappen. Daarna keerde hij terug naar Brünn, waar hij natuurkunde doceerde aan een technische school.



Figuur 8.1:
Mendel (1822–1884)

Tussen 1856 en 1863 voerde Mendel een reeks experimenten uit op een

groot aantal planten in de tuin van zijn abdij. In 1865 stelde hij zijn resultaten voor op twee bijeenkomsten van het Natuurhistorisch Genootschap van Brünn, waarvan hij lid was. Zijn werk, *Experimenten met plantenhybridisatie*, werd het jaar daarop in het Duits gepubliceerd in de handelingen van het Genootschap. Mendel legde uit hoe hij ertoe gekomen was de variaties van erwten te bestuderen, planten die zich op natuurlijke wijze voortplanten door zelfbevruchting en waarvan de zaden verschillende gemakkelijk herkenbare vormen kunnen aannemen: rond of gerimpeld, geel of groen, enz. Door kruising van een plant met ronde zaden met een plant met gerimpelde zaden, merkte hij dat hij altijd hybriden verkreeg die ronde zaden gaven. Hij noemde het karakter „ronde zaden” dominant en het karakter „gerimpelde zaden” recessief. Op dezelfde manier toonde hij aan dat het kenmerk „gele zaden” dominant was en het kenmerk „groene zaden” recessief.

Mendel merkte vervolgens op dat de zelfbevruchting van planten die gekweekt werden uit hybride zaden, in de eerste generatie nieuwe zaden voortbracht die ofwel het dominante ofwel het recessieve kenmerk vertoonden in schijnbaar willekeurige verhoudingen. Bovendien merkte hij op dat door het experiment vele malen te herhalen, hij gemiddeld ongeveer driemaal meer zaden met het dominante dan met het recessieve kenmerk verkreeg. Zo verkreeg hij bij een eerste experiment in totaal 5.474 ronde en 1.850 gerimpelde zaden, wat overeenkomt met een verhouding van 2,96 tot 1. Een tweede experiment leverde een totaal van 6.022 gele en 2.001 groene zaden op, wat overeenkomt met een verhouding¹ van 3,01 tot 1.

Mendel merkte ook op dat van de planten die gekweekt werden uit de zaden van de eerste generatie met het dominante kenmerk, de planten die door zelfbevruchting zaden voortbrachten met hetzij het dominante, hetzij het recessieve kenmerk, ongeveer dubbel zo talrijk waren als de planten die zaden voortbrachten met louter het dominante kenmerk. Bijvoorbeeld, van de 565 planten die opgekweekt werden uit ronde zaden van de eerste generatie, brachten er 372 zowel ronde als gerimpelde zaden voort, terwijl 193 planten enkel ronde zaden voortbrachten; een verhouding van 1,93. Evenzo brachten van de 519 planten die werden gekweekt uit gele zaden van de eerste generatie, 353 planten zowel gele als groene zaden voort, terwijl 166 planten alleen gele zaden voortbrachten; een verhouding van 2,13.

Om deze resultaten te verklaren, kwam Mendel op het briljante idee om

¹Zoals R. A. Fisher (zie hoofdstuk 14) later opmerkte, is de waarschijnlijkheid om experimentele resultaten te verkrijgen die zo dicht bij de theoretische waarde liggen, vrij klein. Waarschijnlijk heeft Mendel zijn gegevens „gemasseerd”. Bijvoorbeeld, in het tweede experiment met $n = 6.022 + 2.001 = 8.023$ zaden, is de kans dat de verhouding minder dan 0,01 van 3 verschilt slechts ongeveer 10 %.

het schijnbare karakter van een zaadje te beschouwen als het resultaat van de associatie van twee verborgen factoren, waarbij elk van deze factoren ofwel dominant (geschreven A) ofwel recessief (geschreven a) is. Er zijn dus drie mogelijke combinaties: AA , Aa en aa . De zaden met de factoren AA of Aa hebben hetzelfde dominante kenmerk A . De zaden met de factoren aa hebben het recessieve kenmerk a . Mendel veronderstelde bovendien dat tijdens de bevruchting de stuifmeelkorrels en de eicellen (de gameten) slechts één van de twee factoren overdragen, elk met een kans $1/2$.

De kruising van zuivere lijnen AA en aa levert dus hybriden op die allemaal de factoren Aa en het dominante kenmerk A hebben. De gameten van de hybride Aa brengen de factor A over met kans $1/2$ en de factor a met kans $1/2$. Zelfbevruchting van een plant gekweekt uit een hybride zaad Aa geeft dus AA met waarschijnlijkheid $1/4$, Aa met waarschijnlijkheid $1/2$ en aa met waarschijnlijkheid $1/4$, zoals weergegeven in Tabel 8.1.

Tabel 8.1: Mogelijke resultaten van de zelfbevruchting van een hybride Aa en hun waarschijnlijkheid in functie van de factoren overgedragen door de mannelijke gameten (rijen) en door de vrouwelijke gameten (kolommen).

Factor (Waarschijnlijkheid)	A ($1/2$)	a ($1/2$)
A ($1/2$)	AA ($1/4$)	Aa ($1/4$)
a ($1/2$)	Aa ($1/4$)	aa ($1/4$)

Mendel merkte op dat de verhoudingen $AA : Aa : aa$, die $1 : 2 : 1$ bedroegen, ook konden worden verkregen door de formele berekening $(A + a)^2 = AA + 2Aa + aa$. Aangezien de zaden AA en Aa het kenmerk A hebben terwijl alleen de zaden aa het kenmerk a hebben, zijn er inderdaad driemaal meer zaden met kenmerk A dan met het kenmerk a . Bovendien zijn er gemiddeld twee keer zoveel zaden Aa als AA . Zelfbevruchting van planten uit de eerste categorie levert zaden op met ofwel het dominante kenmerk (AA of Aa) ofwel het recessieve kenmerk (aa). Zelfbevruchting van planten uit zaad AA brengt altijd zaad AA voort met het dominante kenmerk. Alle waarnemingen zijn dus verklaard.

Mendel keek ook naar de volgende generaties. Uitgaande van N hybride zaden Aa en voor de eenvoud veronderstellende dat elke plant door zelfbevruchting slechts vier nieuwe zaden geeft, berekende hij dat het gemiddelde aantal zaden $(AA)_n$, $(Aa)_n$ en $(aa)_n$ in generatie n zou worden gegeven door Tabel 8.2, waarbij voor de duidelijkheid van de presentatie de resultaten door N zijn gedeeld. Deze getallen worden simpelweg verkregen uit de formules

$$(AA)_{n+1} = (Aa)_n + 4(AA)_n, \quad (8.1)$$

$$(Aa)_{n+1} = 2(Aa)_n, \quad (8.2)$$

$(aa)_{n+1} = (Aa)_n + 4(aa)_n$, die zeggen dat AA na zelfbevruchting vier zaden AA voortbrengt, dat aa vier zaden aa voortbrengt en dat Aa gemiddeld één zaad AA , twee zaden Aa en één zaad aa voortbrengt. Mendel merkte verder op dat $(AA)_n = (aa)_n = 2^{n-1}(2^n - 1)$ en $(Aa)_n = 2^n$.

Tabel 8.2: Opeenvolgende generaties.

n	0	1	2	3	4	5
$(AA)_n$	0	1	6	28	120	496
$(Aa)_n$	1	2	4	8	16	32
$(aa)_n$	0	1	6	28	120	496
totaal	1	4	16	64	256	1.024

Uit vergelijking (8.2) en uit de beginvoorwaarde $(Aa)_0 = 1$ volgt inderdaad dat $(Aa)_n = 2^n$. Door dit te vervangen in vergelijking (8.1), krijgen we dat $(AA)_{n+1} = 4(AA)_n + 2^n$. We realiseren ons gemakkelijk dat $(AA)_n = c2^n$ een bijzondere oplossing is als $c = -1/2$. De algemene oplossing van de „homogene” vergelijking $(AA)_{n+1} = 4(AA)_n$ is $(AA)_n = C4^n$. Als we tenslotte deze twee oplossingen optellen, zien we dat $(AA)_n = C4^n - 2^{n-1}$ voldoet aan de beginvoorwaarde $(AA)_0 = 0$ als $C = 1/2$. Wat de rij $(aa)_n$ betreft, deze voldoet aan hetzelfde recurrentieverband en dezelfde beginvoorwaarde als $(AA)_n$. Dus $(aa)_n = (AA)_n$.

Het aandeel van de hybriden Aa in de totale populatie, die $2^n/4^n = 1/2^n$ bedraagt, wordt bij elke generatie door zelfbevruchting door twee gedeeld.

Het werk van Mendel bleef tijdens zijn leven volledig onopgemerkt. Enkele jaren later probeerde Mendel ook soortgelijke experimenten uit met andere plantensoorten, publiceerde hij enkele artikelen over meteorologie en onderzocht hij de erfelijkheid van bijen. Nadat hij in 1868 abt was geworden, besteedde hij het grootste deel van zijn tijd aan het oplossen van administratieve problemen. Hij overleed in 1884.

Pas in 1900 werd het werk van Mendel uiteindelijk onafhankelijk en bijna gelijktijdig herontdekt door Hugo De Vries in Amsterdam, Carl Correns in Tübingen en Erich von Tschermak in Wenen. Dit zou het begin zijn van een nieuw tijdperk in wat wij nu genetica noemen.

Verdere lectuur

1. Bateson, W.: *Mendel's Principles of Heredity* (1913). archive.org
2. Mendel, J.G.: *Versuche über Pflanzenhybriden* (1866). www.esp.org
3. Fisher, R.A.: Has Mendel's work been rediscovered? *Ann. Sci.* 1, 115–137 (1936). library.adelaide.edu.au

Hoofdstuk 9

Galton, Watson en het probleem van uitsterven (1873–1875)

In 1873 bogen de Britse statisticus Galton en zijn landgenoot, de wiskundige Watson, zich over het probleem van het uitsterven van familienamen, zonder op de hoogte te zijn van het werk van Bienaymé. Watson merkte op dat de kansgenererende functie van het aantal mannen in elke generatie recursief kon worden berekend. Maar hij analyseerde de kans op uitsterven onjuist.

Francis Galton werd geboren in 1822, hetzelfde jaar als Mendel, nabij Birmingham in Engeland. Hij was de jongste van zeven kinderen. Zijn vader was een rijke bankier. Van moederskant was hij de neef van Charles Darwin. Galton begon in 1838 met een studie medicijnen, eerst in een ziekenhuis in Birmingham en later in Londen. In de zomer van 1840 maakte hij zijn eerste lange reis door Europa, tot aan Istanbul. Daarna studeerde hij vier jaar aan het *Trinity College* van de universiteit van Cambridge. Maar zijn vader stierf in 1844 en liet een aanzienlijk fortuin na. Galton gaf het idee op om arts te worden. Hij reisde naar Egypte, Soedan en Syrië. Gedurende de volgende jaren hield hij er een ontspannen levensstijl op na en besteedde hij zijn tijd aan jagen, ballon- en bootreizen en pogingen om de elektrische telegraaf te verbeteren. In 1850 zette hij een verkenningsexpeditie op naar Zuidwest-Afrika (het huidige Namibië). Bij zijn terugkeer naar Engeland in 1852 werd hij verkozen tot lid van de *Royal Geographical Society*. Daar kon hij het nieuws volgen van de expedities naar Oost-Afrika op zoek naar de bron van de Nijl. Hij vestigde zich in Londen en schreef een gids voor reizigers die een bestseller werd. In 1856 werd hij verkozen tot lid van de *Royal Society*. Hij was toen geïnteresseerd in meteorologie en vond het woord „anticyclon” uit. Na de publicatie in 1859 door zijn neef Darwin van *De oorsprong der soorten*, richtte Galton zich op de studie van erfelijkheid. Hij publiceerde *Erfelijk Genie* in 1869, waarin hij beweerde dat intellectuele vermogens door erfelijkheid konden worden overgedragen.

In 1873 publiceerde Alphonse de Candolle, een Zwitserse botanicus, een boek getiteld *Geschiedenis van de wetenschap en van de wetenschappers in de laatste twee eeuwen*, dat ook een essay bevatte over *De respectievelijke*



Figuur 9.1: Galton (links) en Watson (rechts).

invloed van erfelijkheid, variabiliteit en selectie op de ontwikkeling van de menselijke soort en op de waarschijnlijke toekomst van deze soort. Daarin maakte hij de volgende opmerkingen:

„Tussen de nauwkeurige informatie en zeer verstandige meningen van de heren Benoiston de Châteauneuf, Galton en andere statistici, heb ik niet de belangrijke opmerking gezien die zij hadden moeten maken over het *onvermijdelijke* uitsterven van familienamen. Natuurlijk moet elke naam uitsterven [...] Een wiskundige zou kunnen berekenen hoe de afname van de namen of titels zou gebeuren, wetende hoe groot de kans is dat een paar vrouwelijke of mannelijke kinderen krijgt en hoe groot de kans is dat een paar geen kinderen krijgt.”

Het is hetzelfde probleem dat Bienaymé in 1845 had bestudeerd. Maar Candolle, die niet op de hoogte was van het werk van Bienaymé, dacht dat alle families tot uitsterven gedoemd waren. Galton merkte de bovenstaande paragraaf op in Candolle's boek. Omdat hij ook niet op de hoogte was van Bienaymé's werk, presenteerde Galton het als een open probleem aan de lezers van de *Educational Times*:

„Opgave 4.001: Een groot volk, waarvan we alleen de volwassen mannen beschouwen, met N in aantal, en die elk een eigen achternaam dragen, koloniseert een gebied. Hun bevolkingsgroei is zodanig dat in elke generatie a_0 procent van de volwassen mannen geen mannelijke kinderen hebben die volwassen worden; a_1

hebben één zo'n mannelijk kind; a_2 hebben er twee; en zo verder tot a_5 die er vijf hebben.

Bepaal (1) welk deel van hun familienamen zal zijn uitgestorven na r generaties; en (2) hoeveel gevallen er zullen zijn van de familienaam die wordt gehouden door m personen."

Merk op dat het tweede deel van het probleem niet door Bienaymé was behandeld. Galton kreeg geen bevredigend antwoord van de lezers van het tijdschrift en kon blijkbaar zelf de oplossing van het probleem niet vinden. Daarom vroeg hij zijn vriend Henry William Watson, een wiskundige, om te proberen het op te lossen.

Watson werd in 1827 in Londen geboren. Zijn vader was officier bij de Britse marine. Hij studeerde eerst aan het *King's College* in Londen en wijdde zich vervolgens aan de wiskunde aan het *Trinity College*, Universiteit van Cambridge, van 1846 tot 1850, enkele jaren na Galton. Hij werd achtereenvolgens fellow van het *Trinity College*, assistent-meester aan de *City of London School*, docent wiskunde aan het *King's College* en hoogleraar wiskunde aan de *Harrow School* tussen 1857 en 1865. Hij hield van alpinisme en maakte deel uit van een expeditie die in 1855 de top van de Monte Rosa in Zwitserland bereikte. In 1856 werd hij tot diaken gewijd en twee jaar later tot Anglicaans priester. Van 1865 tot zijn emeritaat was hij rector van Berkswell en Barton bij Coventry, een functie die hem voldoende tijd liet voor onderzoek.

Galton en Watson schreven samen een artikel getiteld *Over de waarschijnlijkheid van het uitsterven van families*, dat in 1875 werd gepubliceerd in het *Journal of the Royal Anthropological Institute*. Galton stelde het probleem voor en Watson legde zijn berekeningen en de conclusies uit. Zij gingen ervan uit dat mannen ten hoogste q zonen hebben, waarbij p_k de kans is om k zonen te hebben ($k = 0, 1, 2, \dots, q$). Met andere woorden, $p_k = a_k/100$ als we Galton's oorspronkelijke notaties gebruiken. Dus $p_0 + p_1 + \dots + p_q = 1$. Beschouw de situatie waarin generatie 0 uit één man bestaat. Generatie 1 bestaat uit s mannen met een waarschijnlijkheid p_s . Met behulp van een truc die in zijn tijd bekend was en lang tevoren was ingevoerd door Abraham de Moivre, beschouwde Watson de genererende functie

$$f(x) = p_0 + p_1x + p_2x^2 + \dots + p_qx^q \quad (9.1)$$

geassocieerd met de kansen p_0, \dots, p_q . Laat op dezelfde manier $f_n(x)$ de polynoom zijn waarvan de coëfficiënt van x^s de kans is dat er s mannen zijn in generatie n uitgaande van één man in generatie 0. Dan is $f_1(x) = f(x)$. Watson merkte op dat

$$f_n(x) = f_{n-1}(f(x)), \quad (9.2)$$

een formule die het mogelijk maakt $f_n(x)$ recursief te berekenen.

Stel namelijk

$$f_n(x) = p_{0,n} + p_{1,n}x + p_{2,n}x^2 + \cdots + p_{q^n,n}x^{(q^n)}.$$

Merk op dat er in generatie n ten hoogste q^n mannen zijn. Als er in generatie $n-1$ s mannen zijn, genummerd van 1 tot en met s , noem dan t_1, \dots, t_s het aantal van hun mannelijke nakomelingen. In dat geval zullen er in generatie n t mannen zijn met kans

$$\sum_{t_1 + \cdots + t_s = t} p_{t_1} \times \cdots \times p_{t_s}.$$

Wanneer $s = 0$, is deze kans gelijk aan 1 als $t = 0$ en gelijk aan 0 als $t = 1$. Daarom

$$p_{t,n} = \sum_{s \geq 0} p_{s,n-1} \times \sum_{t_1 + \cdots + t_s = t} p_{t_1} \times \cdots \times p_{t_s}.$$

Hieruit volgt dat

$$\begin{aligned} f_n(x) &= \sum_{t \geq 0} p_{t,n} x^t \\ &= \sum_{s \geq 0} p_{s,n-1} \sum_{t \geq 0} \sum_{t_1 + \cdots + t_s = t} (p_{t_1} x^{t_1}) \times \cdots \times (p_{t_s} x^{t_s}) \\ &= \sum_{s \geq 0} p_{s,n-1} [p_0 x^0 + p_1 x^1 + p_2 x^2 + \cdots]^s \\ &= \sum_{s \geq 0} p_{s,n-1} [f(x)]^s = f_{n-1}(f(x)). \end{aligned}$$

In het bijzonder is de kans x_n op het uitsterven van de familienaam binnen n generaties gelijk aan $p_{0,n}$, wat hetzelfde is als $f_n(0)$. Als eerste voorbeeld nam Watson

$$f(x) = (1 + x + x^2) / 3,$$

d.w.z. $q = 3$ en $p_0 = p_1 = p_2 = 1/3$. Hij berekende de polynomen $f_n(x)$ voor $n = 1, 4$ met behulp van vergelijking (9.2). Hij vond bijvoorbeeld

$$f_2(x) = \frac{1}{3} \left[1 + \frac{1+x+x^2}{3} + \left(\frac{1+x+x^2}{3} \right)^2 \right] = \frac{13 + 5x + 6x^2 + 2x^3 + x^4}{27}$$

en $f_2(0) = 13/27 \approx 0,481$. De berekening van $f_n(x)$ voor $n \geq 3$ wordt erg vervelend, zo vervelend dat Watson al een fout maakte voor $n = 4$. Aangezien

$x_5 = f_5(0) = f_4(f(0))$ de berekening van $f_5(x)$ kan vermijden, kreeg hij de volgende lijst van uitdovingskansen $x_n = f_n(0)$:

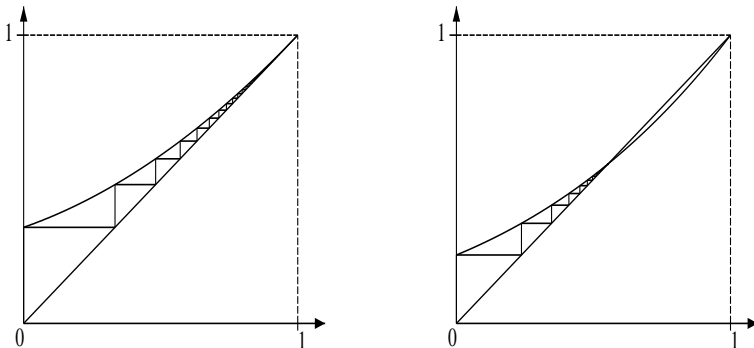
$$x_1 \approx 0,333, \quad x_2 \approx 0,481, \quad x_3 \approx 0,571, \quad x_4 \approx 0,641, \quad x_5 \approx 0,675.$$

De juiste waarden zijn $x_4 = 0,632$ en $x_5 = 0,677$, zoals kan worden nagegaan met de eenvoudige formule $x_n = f(x_{n-1})$ afgeleid door Bienaymé. Zoals we in hoofdstuk 17 zullen zien, kan deze laatste formule ook worden afgeleid uit vergelijking (9.2).

Watson merkte op dat elke man gemiddeld

$$\mathcal{R}_0 = p_1 + 2p_2 + \cdots + qp_q$$

zonen heeft en dat $\mathcal{R}_0 = 1$ in zijn eerste voorbeeld. Men zou dus kunnen denken dat als het aanvankelijke aantal mannelijke familieleden groot genoeg was, de gezinsgrootte ruwweg constant zou blijven. Niettemin beweert Watson dat de uitstervingskans x_n naar 1 convergeert als $n \rightarrow +\infty$, zij het vrij langzaam. Met andere woorden, de hele familie zal uitsterven zoals Candolle had gesuggereerd. De figuur 9.2a, die niet in het oorspronkelijke artikel is opgenomen, en de resultaten van Bienaymé bevestigen dat deze conclusie voor het eerste voorbeeld juist is.



Figuur 9.2: Grafiek van de functies $y = f(x)$ en $y = x$. De uitstervingskans $x_n = f(x_{n-1})$ binnen n generaties is de hoogte van de n -de „trede van de trap”. Links: $f(x) = (1+x+x^2)/3$. Rechts: $f(x) = (3+x)^5/4^5$.

Als tweede voorbeeld beschouwde Watson de binomiale kansverdeling

$$p_k = \binom{q}{k} \frac{a^{q-k} b^k}{(a+b)^q}, \quad (9.3)$$

waarvoor de genererende functie

$$f(x) = \frac{(a+bx)^q}{(a+b)^q}$$

is. Hij berekende $f_2(x)$ en $x_2 = f_2(0)$. Op dat moment realiseerde hij zich dat $x_2 = f(x_1)$ en dat $x_n = f(x_{n-1})$ voor alle n . Maar hij dacht dat deze formule alleen waar was voor het speciale binomiale geval (9.3). Als hij ze toepast op het geval waarin $q = 5$, $a = 3$ en $b = 1$, verkrijgt hij

$$x_1 \approx 0,237, x_2 \approx 0,347, x_3 \approx 0,410, \dots x_9 \approx 0,527, x_{10} \approx 0,533 \dots$$

Watson realiseerde zich dat x_n moet convergeren naar een limiet x_∞ naarmate $n \rightarrow +\infty$ convergeert, die voldoet aan $x_\infty = f(x_\infty) = (a+bx_\infty)^q/(a+b)^q$. Hij merkte op dat $x = 1$ een oplossing van deze vergelijking is, maar realiseerde zich niet dat er andere oplossingen konden zijn wanneer $\mathcal{R}_0 > 1$. Hij concludeerde dus ten onrechte, misleid door Candolle, dat er in alle gevallen sprake is van uitsterving ($x_\infty = 1$), ook in het numerieke voorbeeld dat hij zojuist had bekeken. Figuur 9.2b laat zien dat dit niet het geval is!

Watson merkte op dat het gemiddelde aantal zonen in dit numerieke voorbeeld groter was dan 1 (men kan aantonen dat $\mathcal{R}_0 = qb/(a+b) = 5/4$), wat betekent dat de bevolking de neiging heeft exponentieel toe te nemen. Maar dit hielp hem niet zijn fout te ontdekken. Hij vermoedde zelfs dat het uitsterven van de familienaam zeker was voor elke kansverdeling (p_k), dus niet alleen voor het binomiale geval. We zullen in de hoofdstukken over Haldane en Erlang op dit probleem terugkomen.

Galton zette zijn statistische studie van families voort met een boek, getiteld *Engelse mannen van de wetenschap, hun aard en opvoeding*, dat zich richtte op de genealogie van de leden van de *Royal Society*. Hij raakte ook geïnteresseerd in anthropometrie, de meting van het menselijk lichaam. Hij maakte gebruik van een internationale tentoonstelling in 1884 in Londen om gegevens te verzamelen over een groot aantal mensen. Zijn resultaten werden in 1889 gepubliceerd in een boek getiteld *Natuurlijke erfelijkheid*, waarvan de bijlage het artikel weergaf dat hij in samenwerking met Watson had geschreven. Dit boek introduceerde ook een aantal nieuwe statistische termen zoals „percentiel” en „kwartiel”, evenals het woord „eugenetica”, d.w.z. de

verbetering van de menselijke soort vanuit het oogpunt van erfelijke eigenschappen. Na 1888 ontwikkelde Galton de techniek van het herkennen van vingerafdrukken die enkele jaren later door de Britse politie zou worden gebruikt. Hij bleef ook de respectievelijke rol bestuderen van de erfelijkheid en van het milieu op de fysieke en intellectuele kenmerken van tweelingen, op de grootte van erwten die over verschillende generaties werden gekweekt en op de kleur van muizen die in een laboratorium werden gekweekt. Dit leidde hem tot het begrip „correlatiecoëfficiënt” tussen twee variabelen. In 1904 werd het Galton-laboratorium opgericht in het *University College* in Londen. Galton werd in 1909 geridderd en overleed in 1911.

Watson publiceerde verschillende boeken, met name een verhandeling over de kinetische theorie van gassen in 1876 en een verhandeling over de mathematische theorie van elektriciteit en magnetisme in twee delen (1885 en 1889). Hij werd in 1881 tot lid van de *Royal Society* gekozen en overleed in 1903 te Brighton.

In 1924, in het tweede deel van zijn biografie van Galton, vatte Karl Pearson het artikel over het uitsterven van familienamen samen zonder de fout op te merken. Deze fout zou uiteindelijk in 1930 worden opgemerkt (zie hoofdstuk 18).

Verdere lectuur

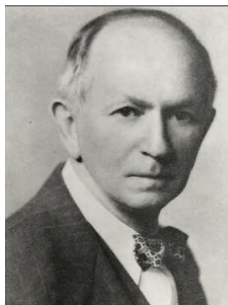
1. De Candolle, A.: *Histoire des sciences et des savants depuis deux siècles*. Georg, Genève (1873). archive.org
2. Galton, F.: *Natural Inheritance*. Macmillan, London (1889). galton.org
3. Galton, F.: *Memories of my Life*. Methuen & Co., London (1908). galton.org
4. Kendall, D.G.: Branching processes since 1873. *J. Lond. Math. Soc.* 41, 385–406 (1966)
5. Pearson, K.: *The Life, Letters and Labours of Francis Galton*, vol. 1/2. Cambridge University Press (1914/1924). galton.org
6. S.H.B.: Henry William Watson, 1827-1903. *Proc. R. Soc. Lond.* 75, 266–269 (1905). gallica.bnf.fr
7. Watson, H.W., Galton, F.: On the probability of the extinction of families. *J. Anthropol. Inst.* 4, 138–144 (1875). galton.org

Hoofdstuk 10

Lotka en de stabiele bevolkingstheorie (1907–1911)

In 1907 begon de Amerikaanse scheikundige Alfred Lotka de relatie tussen geboortecijfer, leeftijdsspecifieke sterftcijfers en het tempo van de bevolkingsgroei te bestuderen met behulp van een continue-tijdmodel. In 1911 publiceerde hij samen met F. R. Sharpe een ander artikel over hetzelfde onderwerp, waarin ook leeftijdsspecifieke vruchtbaarheidscijfers waren opgenomen. De impliciete vergelijking die het bevolkingsgroeipercentage geeft, wordt vaak de „vergelijking van Lotka” genoemd.

Alfred James Lotka werd in 1880 uit Amerikaanse ouders geboren in Lemberg, dat deel uitmaakte van het Oostenrijks-Hongaarse keizerrijk (nu L'viv in Oekraïne). Hij studeerde eerst in Frankrijk en Duitsland en behaalde in 1901 een bachelor's degree in natuur- en scheikunde aan de Universiteit van Birmingham in Engeland. Daarna verbleef hij een jaar in Leipzig, waar de rol van de thermodynamica in de chemie en de biologie werd benadrukt door Wilhelm Ostwald, die in 1909 de Nobelprijs voor scheikunde zou ontvangen. Lotka vestigde zich in 1902 in New York en begon te werken voor de *General Chemical Company*.



Figuur 10.1:
Lotka (1880–1949)

In 1907 en 1911¹ begon Lotka met de studie van de dynamiek van leeftijdsgestructureerde populaties zonder op de hoogte te zijn van het werk van Euler over hetzelfde onderwerp (zie hoofdstuk 3). In tegenstelling tot Euler

¹Het tweede artikel werd geschreven in samenwerking met F. R. Sharpe, een wiskundige van Cornell-universiteit.

ging hij ervan uit dat tijd en leeftijd continue variabelen zijn. Zij $B(t)$ het aantal mannelijke geboorten per tijdseenheid op tijdstip t , $p(x)$ de kans om op leeftijd x nog in leven te zijn en $h(x)$ de vruchtbaarheid op leeftijd x : $h(x)dx$ de kans voor een man om tussen leeftijd x en $x + dx$ één pasgeboren zoon te krijgen als dx oneindig klein is. Dan is $\int_0^{+\infty} p(x)dx$ de levensverwachting bij de geboorte. Bovendien is $B(t-x)p(x)dx$ het aantal mannetjes geboren tussen tijdstip $t-x$ en $t-x+dx$, die op tijdstip t nog in leven zijn. Deze mannetjes hebben $B(t-x)p(x)h(x)dx$ zonen per tijdseenheid op tijdstip t . Het totale aantal mannelijke geboorten op tijdstip t is dus

$$B(t) = \int_0^{+\infty} B(t-x)p(x)h(x)dx.$$

Op zoek naar een exponentiële oplossing voor deze integraalvergelijking in de onbekende $B(t)$ van de vorm $B(t) = be^{rt}$, verkreeg Lotka door beide zijden te delen door $B(t)$ de vergelijking

$$1 = \int_0^{+\infty} e^{-rx}p(x)h(x)dx, \quad (10.1)$$

die nu door demografen „Lotka's vergelijking” wordt genoemd². Euler had de analoge impliciete vergelijking (3.1) verkregen voor de groeisnelheid wanneer tijd en leeftijd discrete variabelen zijn. Lotka merkte op dat het rechterlid van (10.1) een afnemende functie van r is die naar $+\infty$ neigt wanneer $r \rightarrow -\infty$ en naar 0 neigt wanneer $r \rightarrow +\infty$. Er is dus een unieke waarde van r , noem die r^* , zo dat vergelijking (10.1) geldt. Bovendien is $r^* > 0$ als en slechts als

$$\mathcal{R}_0 = \int_0^{+\infty} p(x)h(x)dx > 1. \quad (10.2)$$

De parameter \mathcal{R}_0 (de notatie werd in 1925 door Dublin en Lotka ingevoerd) is het verwachte aantal zonen dat een man tijdens zijn leven kan krijgen.

Lotka stelde dat, ongeacht de initiële leeftijdsstructuur van de bevolking, het aantal mannelijke geboorten per tijdseenheid inderdaad zodanig was dat $B(t) \sim be^{r^*t}$ wanneer $t \rightarrow +\infty$, waarbij b een constante is³. De totale bevolking wordt dan gegeven door $P(t) = \int_0^{+\infty} B(t-x)p(x)dx$. Hieruit volgt dat $P(t)$ ook toeneemt of afneemt zoals e^{r^*t} wanneer $t \rightarrow +\infty$: de groeisnelheid is

²R. A. Fisher kwam onafhankelijk tot dezelfde vergelijking in 1927 en interpreteerde later de wortel r^* als een maatstaf voor „Darwinistische fitness” in de theorie van evolutie door natuurlijke selectie.

³Dit werd in 1941 op exacte wijze bewezen door Willy Feller, die toen hoogleraar wiskunde was aan de Brown-universiteit in de Verenigde Staten. Een probabilistische benadering werd in 1968 ontwikkeld door Crump, Mode en Jagers.

gelijk aan r^* . Bovendien neigt de leeftijdsstructuur van de bevolking, gegeven door $B(t-x)p(x)/P(t)$, naar $\frac{e^{-r^*x}p(x)}{\int_0^{+\infty} e^{-r^*y}p(y)dy}$. Dit is wat Lotka een „stabile bevolking” noemde: de leeftijds piramide behoudt door de tijd heen dezelfde vorm, maar de totale bevolking neemt exponentieel toe of af. De conclusie is dus dezelfde als in Euler’s discrete-tijd model. Maar Lotka’s studie houdt rekening met de leeftijdsafhankelijkheid van de vruchtbaarheid. Zij is dus in zekere zin algemener dan die van Euler.

Lotka bleef zijn hele leven aan dit onderwerp werken. In 1908–1909 hervatte hij zijn studie aan Cornell-universiteit om een mastergraad te behalen. Hij werkte voor het *National Bureau of Standards* van 1909 tot 1911 en als redacteur van het tijdschrift *The Scientific American Supplement* van 1911 tot 1914. In 1912 promoveerde hij aan de Universiteit van Birmingham door de artikelen te verzamelen die hij sinds 1907 had gepubliceerd over populatiedynamica en demografie. Tijdens de Eerste Wereldoorlog werkte hij opnieuw voor de *General Chemical Company* aan de manier om stikstof uit de atmosfeer vast te leggen. In 1920 maakte een van zijn artikelen over biologische oscillaties (zie hoofdstuk 13) diepe indruk op Raymond Pearl, een professor in de biometrie aan de Johns Hopkins-universiteit die net de logistische vergelijking had „herontdekt” (zie hoofdstuk 6). In de hoop een baan te vinden aan het Rockefeller Instituut voor Medisch Onderzoek in New York, werkte Lotka aan de wiskundige modellen die Ross had ontwikkeld voor malaria (zie hoofdstuk 12). Uiteindelijk kreeg hij een tweejarige beurs van de Johns Hopkins-universiteit, die hem in staat stelde een boek te schrijven, getiteld *Elementen van de fysische biologie*, dat in 1925 werd gepubliceerd. Daarna werd hij hoofd van de onderzoeksafdeling van de *Metropolitan Life Insurance Company* in New York. Hij legde zich toe op de wiskundige analyse van demografische vraagstukken en publiceerde verschillende boeken in samenwerking met een collega, de statisticus en vice-president van de maatschappij Louis Israel Dublin: *De geldwaarde van een man* (1930), *De lengte van het leven* (1936) en *Vijfentwintig jaar vooruitgang in de gezondheid* (1937). Hij werd in 1938-1939 gekozen tot voorzitter van de *Population Association of America*. Onder zijn verschillende statistische studies stelt „Lotka’s wet” (die teruggaat tot 1926) dat het aantal auteurs dat n artikelen heeft geschreven op een bepaald wetenschappelijk gebied min of meer afneemt als $1/n^2$ naarmate n toeneemt.

Lotka publiceerde ook een boek in het Frans, getiteld *De Analytische Theorie van Biologische Associaties*. Het eerste deel, dat meer filosofisch van aard was, verscheen in 1934. Het tweede, meer technische deel, verscheen in 1939 en bevatte een samenvatting van al zijn onderzoek op het gebied

van de menselijke demografie sinds 1907. In zijn boek presenteerde Lotka ook zijn bijdrage aan het probleem van het uitsterven van familienamen. Na de publicatie in 1930 van het eerste artikel van Steffensen over dit onderwerp (zie hoofdstuk 18), had hij de theorie toegepast op de gegevens van de volkstelling van 1920 van de blanke bevolking van de Verenigde Staten. Hij stelde vast dat de waargenomen verdeling $(p_k)_{k \geq 0}$ van het aantal zonen goed benaderd wordt door een afnemende meetkundige wet voor alle $k \geq 1$: $p_0 = a$, $p_k = bc^{k-1}$ ($k \geq 1$), met $a = 0,4825$, $b = 0,2126$ en $c = 1 - b/(1 - a)$. Op deze manier is $\sum_{k \geq 0} p_k = 1$. De bijbehorende genererende functie is $f(x) = a + b \sum_{k=1}^{+\infty} c^{k-1} x^k = a + \frac{bx}{1-cx}$. De twee oplossingen van de vergelijking $x = f(x)$ zijn $x = 1$ en $x = a/c$. De uitdovingskans x_∞ is de kleinste van deze twee oplossingen (zie hoofdstuk 7). Met de numerieke waarden voor de Verenigde Staten vond hij $x_\infty \approx 0,819$, terwijl het gemiddeld aantal zonen $\mathcal{R}_0 = f'(1) = (1 - a)^2/b \approx 1,260$ was. Ondanks een gemiddeld aantal kinderen (zonen en dochters inbegrepen) dat dicht bij 2,5 ligt, is de kans op uitsterven van de familienaam groter dan 80 %.

Lotka werd in 1942 verkozen tot voorzitter van de *American Statistical Association*. Hij ging in 1947 met pensioen en overleed in 1949 in New Jersey. Een nieuwe editie van zijn boek uit 1925 verscheen in 1956 onder de iets andere titel *Elementen van wiskundige biologie*.

Verdere lectuur

1. Crump, K.S., Mode, C.J.: A general age-dependent branching process. *J. Math. Anal. Appl.* 24, 494–508 (1968)
2. Dublin, L.I., Lotka, A.J.: On the true rate of natural increase. *J. Amer. Stat. Assoc.* 20, 305–339 (1925)
3. Feller, W.: On the integral equation of renewal theory. *Ann. Math. Stat.* 12, 243–267 (1941). projecteuclid.org
4. Fisher, R.A.: The actuarial treatment of official birth records. *Eugen. Rev.* 19, 103–108 (1927). digital.library.adelaide.edu.au
5. Gridgeman, N.T.: Lotka, Alfred James. In Gillespie, C.C. (ed.) *Dictionary of Scientific Biography*, vol. 8, 512. Scribner, New York (1981)
6. Lotka, A.J.: Relation between birth rates and death rates. *Science* 26, 21–22 (1907) → Smith & Keyfitz (1977).
7. Lotka, A.J.: *Théorie analytique des associations biologiques*, 2^e partie. Hermann, Paris (1939) gallica.bnf.fr
8. Sharpe, F.R., Lotka, A.J.: A problem in age-distribution. *Philos. Mag. Ser. 6*, 21, 435–438 (1911) → Smith & Keyfitz (1977).
9. Smith, D.P., Keyfitz, N.: *Mathematical Demography*. Springer, Berlin (1977)
10. Tanner, A.: *Von Molekülen, Parasiten und Menschen – A. J. Lotka und die Mathematisierung des Lebens*. ETH Zürich (2014) doi:10.3929/ethz-a-010209129

Hoofdstuk 11

De wet van Hardy-Weinberg (1908)

In 1908 ontdekten de Britse wiskundige Hardy en de Duitse arts Weinberg onafhankelijk van elkaar dat in een oneindig grote populatie die zich volgens de wetten van Mendel willekeurig voortplant, de frequenties van de genotypen verkregen uit twee allelen door de generaties heen constant blijven. Hun wiskundig model was een van de uitgangspunten van de populatiegenetica.

Godfrey Harold Hardy werd in 1877 geboren in Surrey, Engeland. Zijn ouders waren onderwijzers. Hij studeerde vanaf 1896 wiskunde aan het *Trinity College* van de universiteit van Cambridge, werd *fellow* van zijn *college* in 1900 en docent wiskunde in 1906. Na een eerste boek over *De integratie van functies van een enkele variabele* (1905), publiceerde hij in 1908 *Een cursus in zuivere wiskunde*, die vele malen werd herdrukt en in vele talen werd vertaald.



Figuur 11.1:
Hardy (1877–1947)

In die tijd had de herontdekking van het werk van Mendel enige twijfels doen rijzen. Sommige biologen vroegen zich af waarom de dominante eigenschappen niet frequenter werden van generatie op generatie. Reginald Punnett, die in 1905 een boek had geschreven getiteld *Mendelisme*, stelde de vraag aan Hardy, met wie hij cricket speelde in Cambridge. Hardy schreef zijn oplossing in een artikel over *Mendeliaanse verhoudingen in een gemengde populatie*, dat in 1908 werd gepubliceerd. Om de analyse te vereenvoudigen,

stelde hij zich de situatie voor van een grote populatie waarbij de keuze van de seksuele partner willekeurig zou zijn. Bovendien beperkte hij zijn aandacht tot slechts twee factoren (of „allelen”) A en a , waarbij A dominant is en a recessief. Laat voor generatie n p_n de frequentie zijn van het „genotype” AA , $2q_n$ die van Aa en r_n die van aa . Uiteraard is $p_n + 2q_n + r_n = 1$. Hardy veronderstelde ook dat geen van deze genotypes tot een overmatige sterfte of een daling van de vruchtbaarheid leidde in vergelijking met de twee andere genotypes. De frequenties in generatie $n + 1$ kunnen gemakkelijk berekend worden door op te merken dat één willekeurig gekozen individu in generatie n allel A overdraagt met een kans $p_n + q_n$: ofwel is het genotype AA en wordt allel A zeker overgedragen, ofwel is het genotype Aa en wordt allel A met 50 % kans overgedragen. Evenzo wordt allel a met een waarschijnlijkheid van $q_n + r_n$ overgedragen. Men kan dus Tabel 11.1 op dezelfde manier samenstellen als Tabel 8.1.

Tabel 11.1: Berekening van de frequenties van de genotypes in generatie $n + 1$ uit de frequenties van de allelen van de ouders (rijen zijn voor de moeder, kolommen voor de vader).

allel frequentie	A $p_n + q_n$	a $q_n + r_n$
A $p_n + q_n$	AA $(p_n + q_n)^2$	Aa $(p_n + q_n)(q_n + r_n)$
a $q_n + r_n$	Aa $(p_n + q_n)(q_n + r_n)$	aa $(q_n + r_n)^2$

De frequenties van de genotypes AA , Aa en aa in generatie $n + 1$ zijn respectievelijk p_{n+1} , $2q_{n+1}$ en r_{n+1} . Hardy vond dus dat

$$p_{n+1} = (p_n + q_n)^2, \quad (11.1)$$

$$2q_{n+1} = 2(p_n + q_n)(q_n + r_n), \quad (11.2)$$

$$r_{n+1} = (q_n + r_n)^2. \quad (11.3)$$

Vervolgens onderzocht hij onder welke voorwaarden de frequenties van de genotypen over de generaties heen constant konden blijven, gelijk aan p , $2q$ en r . Aangezien per definitie $p + 2q + r = 1$, zien we dat vergelijkingen (11.1)-(11.3) allemaal dezelfde voorwaarde opleveren, namelijk $q^2 = pr$.

Bijvoorbeeld, de eerste vergelijking geeft $p = (p + q)^2 = p^2 + 2pq + q^2$, wat gelijk is aan $p(1 - p - 2q) = q^2$ en tenslotte aan $pr = q^2$.

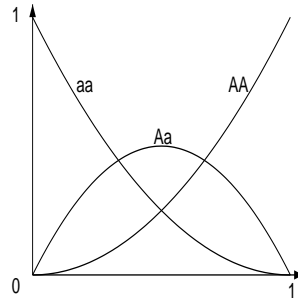
Uitgaande van willekeurige beginvoorwaarden $(p_0, 2q_0, r_0)$ met $p_0 + 2q_0 + r_0 = 1$, merkte Hardy op dat

$$q_1^2 = (p_0 + q_0)^2 (q_0 + r_0)^2 = p_1 r_1.$$

De toestand $(p_1, 2q_1, r_1)$ is dus al een evenwicht. Dus $(p_n, 2q_n, r_n)$ blijft gelijk aan $(p_1, 2q_1, r_1)$ voor alle $n \geq 1$. Als we $x = p_0 + q_0$ stellen voor de frequentie van allel A in generatie 0, dan is $1 - x = q_0 + r_0$ de frequentie van allel a . Als we het systeem (11.1)–(11.3) nog eens gebruiken, krijgen we

$$p_n = x^2, \quad 2q_n = 2x(1-x), \quad r_n = (1-x)^2$$

voor alle $n \geq 1$ (figuur 11.2).



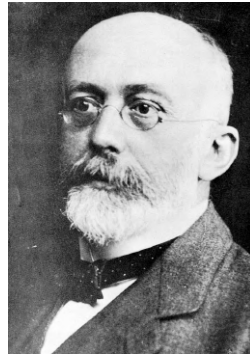
Figuur 11.2: Grafieken van de functies x^2 , $2x(1-x)$ en $(1-x)^2$ die overeenkomen met de evenwichtsfrequenties van de genotypen AA , Aa en aa .

Samenvattend leiden de bovenstaande hypothesen tot een wet die dicteert dat de frequenties van de genotypes AA , Aa en aa onveranderd blijven doorheen de generaties. De theorie van Mendel leidt dus niet tot een progressieve toename van de frequentie van het dominante kenmerk zoals eerst werd gedacht.

Enkele jaren later zou Fisher wijzen op een belangrijk uitvloeisel van deze wet: bij een eerste benadering (d.w.z. uitgaande van de veronderstelling dat de hypothesen van het model realistisch zijn) behoudt een populatie een constante genetische variatie. Deze constatering lost een van de problemen op die Darwins theorie van evolutie door natuurlijke selectie oproep. Darwin dacht namelijk, net als zijn tijdgenoten, dat bij elke generatie de fysiologische kenmerken van de kinderen een soort gemiddelde waren van de kenmerken van de twee ouders, waarbij elke ouder de helft bijdroeg. Dit idee werd later grondig bestudeerd met behulp van statistiek door Francis Galton en zijn opvolger in het biometrisch laboratorium, Karl Pearson. Als dit waar was, zou de

variantie van deze kenmerken in een populatie bij elke generatie moeten halveren en zou er spoedig een zodanige homogeniteit optreden dat natuurlijke selectie, de veronderstelde verklaring van evolutie, onmogelijk zou zijn. Er zouden echter verscheidene jaren nodig zijn om dit middelingsmechanisme te verwerpen, waarbij biometristen het standpunt van Darwin verdedigden en onwillig waren om toe te geven dat de wetten van Mendel onmisbaar zijn om evolutie te begrijpen.

Na dit werk in 1908 keerde Hardy terug naar de zuivere wiskunde. In zijn autobiografie, *Apologie van een wiskundige*, beweerde hij zelfs met trots dat hij ontdekkingen van enig praktisch nut had vermeden. In 1910 werd hij gekozen tot lid van de *Royal Society*. In 1913 ontdekte hij het Indiase wonderkind Ramanujan en nodigde hem uit om in Cambridge te komen werken. Na de Eerste Wereldoorlog werd hij hoogleraar aan de universiteit van Oxford en zette hij de vruchtbare samenwerking met zijn landgenoot Littlewood voort. Tussen 1931 en 1942 was hij opnieuw hoogleraar in Cambridge. Hij publiceerde vele boeken, vaak in samenwerking: *Orders van Oneindigheid* (1910), *De Algemene Theorie van Dirichlet's Reeksen* met Marcel Riesz (1915), *De Ongelijkheden* met Littlewood en Pólya (1934), *Een Inleiding tot de theorie van Getallen* met E. M. Wright (1938), *De Ramanujan* (1940), *Fourierreeksen* met Rogosinski (1944) en *Divergente reeksen* (1949). Hij overleed in Cambridge in 1947.



Figuur 11.3:
Weinberg (1862–1937)

Enkele tientallen jaren later merkte men dat de wet van Hardy voor genfrequenties ook in datzelfde jaar 1908 was ontdekt door een Duitse arts, Wilhelm Weinberg. Weinberg werd in 1862 in Stuttgart geboren. Na studies in Tübingen en München tot aan zijn doctoraat in de geneeskunde, had hij verschillende jaren gewerkt in ziekenhuizen in Berlijn, Wenen en Frankfurt. Hij vestigde zich in 1889 in Stuttgart als huisarts en verloskundige. Ondanks

zijn drukke werkzaamheden had hij toch tijd gevonden om veel artikelen te schrijven in Duitse wetenschappelijke tijdschriften. In 1901 had hij vanuit statistisch oogpunt de frequentie van tweelingen van hetzelfde geslacht bestudeerd. Het artikel van 1908, waarin hij dezelfde wetmatigheid verklaarde als Hardy had gevonden, was gepubliceerd in een plaatselijk wetenschappelijk tijdschrift en was niet opgevallen. Maar in tegenstelling tot Hardy had hij deze studie de volgende jaren voortgezet, waarbij hij bijvoorbeeld de veralgemening had ontdekt voor het geval dat er meer dan twee allelen zijn. Hij had ook een bijdrage geleverd op het gebied van de medische statistiek. Weinberg overleed in 1937. Na de herontdekking van zijn artikel uit 1908 noemden genetici de wet van de stabiliteit van genotype frequenties de „wet van Hardy-Weinberg”.

Tegenwoordig wordt deze wet vaak als volgt gebruikt. Als een zeldzaam recessief allel a geen invloed heeft op overleving of vruchtbaarheid en als we de frequentie x^2 van het genotype aa kennen omdat aa een bepaald fenotype voortbrengt, dan kunnen we x berekenen en de frequentie $2x(1-x) \approx 2x$ van het genotype Aa schatten. Als voorbeeld, als de frequentie van aa $1/20.000$ is, dan krijgen we $x \approx 1/140$. Dus $2x \approx 1/70$ is de frequentie van het genotype Aa . Het recessieve allel a , dat bij inspectie van de fenotypes zeer zeldzaam lijkt, is in feite niet zo zeldzaam.

Verdere lectuur

1. Hardy, G.H.: Mendelian proportions in a mixed population. *Science* 28, 49–50 (1908). esp.org
2. Hardy, G.H.: *A Mathematician's Apology*. Cambridge University Press (1940). archive.org
3. Punnett, R.C.: *Mendelism*, 2nd edn. Cambridge University Press (1907). archive.org
4. Stern, C.: The Hardy–Weinberg law. *Science* 97, 137–138 (1943)
5. Stern, C.: Wilhelm Weinberg 1862–1937. *Genetics* 47, 1–5 (1962)
6. Titchmarsh, E.C.: Godfrey Harold Hardy, 1877–1947. *Obit. Not. Fellows R. Soc.* 6, 446–461 (1949)
7. Weinberg, W.: Über den Nachweis der Vererbung beim Menschen. *Jahresh. Wuertt. Ver. vaterl. Natkd.* 64, 369–382 (1908). biodiversitylibrary.org

Hoofdstuk 12

Ross en malaria (1911)

In 1911 bestudeerde de Britse arts Ronald Ross, die onlangs in 1902 de Nobelprijs had gekregen voor zijn werk op het gebied van malaria, een stelsel differentiaalvergelijkingen dat de verspreiding van deze ziekte modelleert. Hij toonde aan dat malaria alleen kan blijven bestaan als het aantal muggen boven een bepaalde drempelwaarde ligt. Daarom is het niet nodig om alle muggen te doden om malaria uit te roeien – het is voldoende om slechts een bepaalde fractie ervan te doden. Soortgelijke epidemische modellen werden later ontwikkeld door Kermack en McKendrick.

Ronald Ross werd in 1857 geboren in het noorden van India, waar zijn vader officier was in het Britse leger. Hij studeerde geneeskunde in Londen, maar schreef liever gedichten en drama's. Nadat hij één jaar als chirurg op een schip had gewerkt, slaagde hij erin om in 1881 in dienst te treden als legerarts in India. Zijn medisch werk in India liet hem veel vrije tijd, waarin hij literaire werken schreef en zelf wiskunde studeerde. Tijdens een verlof in Engeland in 1888 behaalde hij een diploma in volksgezondheid en studeerde hij bacteriologie, een nieuwe wetenschap die enkele jaren vroeger door Pasteur en Koch was gecreëerd. Bij zijn terugkeer in India begon Ross malaria te bestuderen. Tijdens een tweede verlof in 1894 ontmoette hij in Londen Patrick Manson, een specialist in tropische geneeskunde, die hem onder de microscoop liet zien wat de Franse militaire arts Alphonse Laveran in 1880 had opgemerkt: het bloed van malariapatiënten bevat parasieten. Manson suggereerde dat de parasieten afkomstig konden zijn van muggen omdat hij zelf in China de parasiet van een andere tropische ziekte (filariasis) in deze insecten had ontdekt. Ross geloofde echter dat de mens met de parasiet werd geïnfecteerd door het drinken van water dat door de muggen was besmet. Van 1895 tot 1898 zette hij zijn onderzoek in India voort en testte hij Manson's idee uit. In 1897 ontdekte hij in de maag van een bepaalde muggensoort die hij nog niet had bestudeerd (*anopheles*) enkele parasieten die vergelijkbaar waren met de parasieten die door Laveran eerder werden waargenomen. Omdat zijn superieuren hem naar Calcutta stuurden tijdens een seizoen waarin malaria zelden voorkwam, besloot hij malaria bij kooivogels te bestuderen. Hij vond de parasiet in de speekselklieren van anophelesmuggen en slaagde

erin experimenteel gezonde vogels te infecteren. Om dit te doen liet hij de vogels door muggen bijten: dit toonde aan dat malaria door muggenbeten wordt overgebracht en niet door het drinken van besmet water. In 1899 verliet Ross de Indiase Medische Dienst om les te gaan geven aan de *Liverpool School of Tropical Medicine*, die een jaar eerder was opgericht. Hij werd in 1901 verkozen tot lid van de *Royal Society* en ontving in 1902 de Nobelprijs voor fysiologie of geneeskunde voor zijn werk op het gebied van malaria. Hij reisde naar Afrika, naar Mauritius en in het Middellandse-Zeegebied om er de strijd tegen de muggen te populariseren. De methode was succesvol in Egypte langs het Suezkanaal, langs het in aanbouw zijnde Panamakanaal, in Cuba en in Maleisië. In sommige andere gebieden was het succes minder groot. Ross publiceerde een *Verslag over de preventie van malaria op Mauritius* in 1908 en *De preventie van malaria* in 1910.



Figuur 12.1: Ross (1857–1932)

Ondanks het feit dat hij een bewijs leverde van de rol van bepaalde muggen bij de overdracht van malaria, werd Ross met scepsis ontvangen toen hij beweerde dat malaria uitgeroeid kon worden door eenvoudigweg het aantal muggen te verminderen. In de tweede editie van zijn boek *De preventie van malaria*, gepubliceerd in 1911, probeerde hij wiskundige modellen te bouwen van de overdracht van malaria om zijn bewering te staven. Één van zijn modellen bestond uit een stelsel van twee differentiaalvergelijkingen. Laten we de volgende notaties invoeren:

- N : totale menselijke populatie in een bepaald gebied;
- $I(t)$: aantal met malaria besmette mensen op tijdstip t ;
- n : totale muggenpopulatie (constant verondersteld);

- $i(t)$: aantal muggen besmet met malaria;
- b : bijtfrequentie van muggen;
- p (respectievelijk p'): kans op overdracht van malaria van mens op mug (respectievelijk van mug op mens) tijdens één beet;
- a : snelheid waarmee mensen herstellen van malaria;
- m : muggensterfte.

Gedurende een klein tijdsinterval dt bijt elke besmette mug bdt mensen, waarvan een fractie gelijk aan $\frac{N-I}{N}$ nog niet besmet is. Rekening houdend met de overdrachtskans p' worden er $b p' i \frac{N-I}{N} dt$ nieuwe mensen geïnfecteerd. In hetzelfde tijdsinterval is het aantal mensen dat herstelt $aI dt$. Dus,

$$\frac{dI}{dt} = b p' i \frac{N-I}{N} - aI.$$

Op dezelfde manier bijt elke niet-geïnfecteerde mug bdt mensen, waarvan een fractie gelijk aan I/N al besmet is. Rekening houdend met de overdrachtskans p worden er $b p (n-i) \frac{I}{N} dt$ nieuwe muggen geïnfecteerd. Aangenomen dat de infectie geen invloed heeft op de sterfte, is het aantal muggen dat sterft $mi dt$. Dus,

$$\frac{di}{dt} = b p (n-i) \frac{I}{N} - mi.$$

Aangezien malaria in de meeste besmette landen permanent bestaat, beschouwde Ross enkel de stationaire toestanden van zijn stelsel van twee vergelijkingen: het aantal besmette mensen $I(t)$ en het aantal besmette muggen $i(t)$ blijven constant doorheen de tijd ($dI/dt = 0$ en $di/dt = 0$). Ten eerste is er dus altijd de stationaire toestand met $I = 0$ en $i = 0$, die overeenkomt met een afwezigheid van malaria. En ten tweede zocht Ross naar een stationaire toestand waarbij $I > 0$ en $i > 0$ en vond dat

$$I = N \frac{1 - amN/(b^2 p p' n)}{1 + aN/(b p' n)}, \quad i = n \frac{1 - amN/(b^2 p p' n)}{1 + m/(b p)}. \quad (12.1)$$

Door de stationaire vergelijkingen te delen door het product $I \times i$, wordt het probleem een lineair stelsel van twee vergelijkingen met twee onbekenden $1/I$ en $1/i$,

$$\frac{b p'}{I} - \frac{a}{i} = \frac{b p'}{N}, \quad -\frac{m}{I} + \frac{b p n}{N i} = \frac{b p}{N}.$$

Deze berekening kan gemakkelijk worden uitgevoerd.

Men kan opmerken dat $I > 0$ en $i > 0$ als het aantal muggen boven een kritische drempel ligt:

$$n > n^* = \frac{amN}{b^2 p p'}.$$

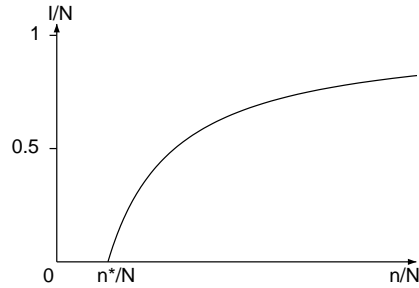
In dit geval komt de stationaire toestand overeen met de situatie waarin de ziekte endemisch is, d.w.z. permanent aanwezig. Ross concludeerde dat als het aantal muggen n tot onder de kritische drempel n^* wordt teruggebracht, de enige overblijvende stationaire toestand $I = 0$ en $i = 0$ is, zodat malaria zou moeten verdwijnen. Met name is het niet nodig alle muggen uit te roeien om malaria uit te roeien. Dit is precies het punt dat Ross met zijn model wilde benadrukken.

Om zijn theorie te illustreren, zocht Ross naar redelijke numerieke waarden voor de parameters van zijn model. Hij nam aan dat:

- de sterfte van muggen zodanig is dat na tien dagen nog slechts een derde in leven is; dus $e^{-10m} = \frac{1}{3}$ en $m = (\log 3)/10$ per dag;
- de helft van de mensen na drie maanden nog besmet is; dus $e^{-90a} = 1/2$ en $a = (\log 2)/90$ per dag;
- één op de acht muggen elke dag bijt; dus $e^{-b} = 1 - 1/8$ en $b = \log(8/7)$ per dag.
- Besmette muggen zijn meestal niet besmettelijk gedurende de eerste tien dagen na hun besmetting omdat de parasieten verschillende stadia van transformatie moeten doorlopen. Aangezien een derde van de muggen tien dagen kan overleven, nam Ross aan dat ook ongeveer een derde van alle geïnfecteerde muggen besmettelijk is: $p' = 1/3$;
- $p = 1/4$.

Ross kon dan met formule (12.1) de besmette fractie I/N in de menselijke populatie berekenen als functie van de verhouding n/N tussen de muggen- en de mensenpopulatie. Zijn resultaten werden in een tabel weergegeven die gelijkwaardig is aan figuur (12.2).

Uit de vorm van de curve blijkt dat de fractie besmette mensen al groter is dan 50% als de verhouding n/N net iets boven de kritische waarde n^*/N ligt. Maar deze fractie verandert niet veel wanneer de verhouding n/N verder toeneemt. Dit verklaart waarom de correlatie tussen het aantal muggen en de aanwezigheid van malaria nooit eerder was opgemerkt. Ross merkte echter op



Figuur 12.2: Fractie I/N van besmette mensen als functie van de verhouding n/N tussen de muggen- en de mensenpopulatie.

dat de numerieke waarde van de drempel n^*/N zeer gevoelig was voor kleine veranderingen in de bijtfrequentie b , maar dat dit de algemene vorm van de curve in figuur 12.2 niet veranderde. Zijn kwalitatieve uitleg is belangrijker dan de kwantitatieve resultaten, die te lijden hebben onder de onzekerheid rond de numerieke waarden van de parameters.

Om de door Ross ontdekte kritische drempel n^* te interpreteren¹, beschouwen we een besmette mens die in een malariavrije mensen- en muggenpopulatie wordt binnengebracht. Deze mens blijft gemiddeld geïnficeerd gedurende een periode gelijk aan $1/a$. Hij of zij krijgt bn/N beten per tijds-eenheid, dus gemiddeld $bn/(aN)$ beten in totaal gedurende de besmetting. Hij of zij infecteert dus gemiddeld $bpn/(aN)$ muggen. Elk van deze besmette muggen leeft gemiddeld gedurende een periode gelijk aan $1/m$, bijt b/m mensen en besmet bp'/m mensen. In totaal, na de overdracht van de eerste besmette mens op de muggen en van deze muggen op andere mensen, is het gemiddelde aantal nieuw besmette mensen het product van de vorige twee resultaten, d.w.z.

$$\mathcal{R}_0 = \frac{b^2 p p' n}{a m N}. \quad (12.2)$$

Dit is dus het aantal secundaire gevallen bij de mens ten gevolge van één primair geval bij de mens. Het besmettingsproces dat continu in de tijd verloopt, kan ook via opeenvolgende generaties worden beschouwd. Malaria kan de populatie alleen binnendringen als $\mathcal{R}_0 > 1$. Deze voorwaarde komt nauwkeurig overeen met $n > n^*$.

Om af te sluiten pleitte Ross voor meer wiskundige modellering in de epidemiologie:

¹Deze interpretatie werd pas lang na Ross's werk benadrukt.

“In feite moet elke epidemiologie, die zich bezighoudt met de variatie van een ziekte in tijd en in plaats, mathematisch worden bekeken. Ongeacht hoeveel variabelen er ook bij betrokken zijn, zal ze helemaal wetenschappelijk kunnen worden bekeken. Zeggen dat een ziekte van bepaalde factoren afhangt, is niet veel zeggen, zolang men niet ook een schatting kan maken van de mate waarin elke factor invloed heeft op het gehele resultaat. Eigenlijk is de methode van wiskundige verwerking niets anders dan de toepassing van een zorgvuldige redenering op de problemen in kwestie.”

Ross werd geadeld in 1911. Hij verhuisde naar Londen en werd adviseur voor het Britse leger tijdens de Eerste Wereldoorlog. In 1923 publiceerde hij zijn autobiografie, *Memoires met een volledig verslag van het grote malariaprobleem en de oplossing ervan*. In 1926 werd het Ross Instituut voor Tropische Ziekten opgericht (nu onderdeel van de *London School of Hygiene and Tropical Medicine*), waarvan hij directeur werd. Ross overleed in Londen in 1932.

Verdere lectuur

1. G.H.F.N.: Sir Ronald Ross, 1857-1932. *Obit. Not. Fellows Roy. Soc.* 1, 108–115 (1933)
2. Ross, R.: *The Prevention of Malaria*, 2nd edn. John Murray, London (1911) archive.org
3. Ross, R.: *Memoirs with a Full Account of the Great Malaria Problem and its Solution*. John Murray, London (1923) archive.org
4. Rowland, J.: *The Mosquito Man, The Story of Sir Ronald Ross*. Roy Publishers, New York (1958)

Hoofdstuk 13

Lotka, Volterra en het predator-prooi systeem (1920–1926)

In 1920 bestudeerde Alfred Lotka een roofdier-prooi model en toonde aan dat de populaties permanent konden oscilleren. Hij werkte deze studie uit in zijn boek *Elementen van de fysische biologie* uit 1925. In 1926 raakte de Italiaanse wiskundige Vito Volterra toevallig geïnteresseerd in hetzelfde model om een vraag van de bioloog Umberto d’Ancona te beantwoorden: waarom werden er meer roofvissen gevangen door de vissers in de Adriatische Zee tijdens de Eerste Wereldoorlog, toen de visserij-inspanning er laag was?

In 1920 publiceerde Lotka een artikel getiteld *Analytische noot over bepaalde ritmische relaties in organische systemen*. Hij was al enige jaren geïnteresseerd in enkele chemische reacties die in laboratoriumexperimenten vreemde voorbijgaande oscillaties vertoonden. Het doel van zijn artikel was te suggereren dat een systeem van twee biologische soorten zelfs permanent zou kunnen oscilleren. Het voorbeeld dat hij beschouwde was dat van een populatie planteneters die zich voeden met planten. Laat, naar analogie met de vergelijkingen die in de chemische kinetiek worden gebruikt, $x(t)$ de totale massa van de planten zijn en $y(t)$ de totale massa van de herbivoren op tijdstip t . Lotka gebruikte als model het volgende stelsel differentiaalvergelijkingen

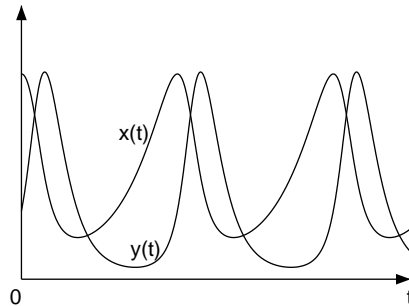
$$\frac{dx}{dt} = ax - bxy, \quad (13.1)$$

$$\frac{dy}{dt} = -cy + dxy, \quad (13.2)$$

waarbij de parameters a , b , c en d alle positief zijn. Parameter a is de groeisnelheid van planten als er geen planteneters zijn, terwijl c de afnamesnelheid van de populatie planteneters is als er geen planten zijn. De termen $-bxy$ en dxy drukken uit dat hoe meer dieren en planten er zijn, hoe groter de massatransfer van planten naar dieren is (de transfer omvat enig massaverlies, dus $d \leq b$). Als men $dx/dt = 0$ en $dy/dt = 0$ instelt, vond Lotka dat er twee stationaire toestanden zijn:

- $(x = 0, y = 0)$, is de populatie van herbivoren uitgestorven en zijn er geen planten meer;
- $(x = c/d, y = a/b)$, herbivoren en planten naast elkaar bestaan.

Hij schreef ook zonder bewijs dat indien op tijdstip $t = 0$, $(x(0), y(0))$ niet een van deze twee stationaire toestanden is, de functies $x(t)$ en $y(t)$ periodiek oscilleren: er is een getal $T > 0$ zodanig dat $x(t+T) = x(t)$ en $y(t+T) = y(t)$ voor alle $t > 0$ (figuur 13.1)¹. Als de planten bijvoorbeeld zeer overvloedig zijn, zal de populatie van herbivoren toenemen, waardoor de totale massa van de planten afneemt. Wanneer deze massa ontoereikend wordt om de plantenters te voeden, zullen sommige dieren van honger sterven en zal de totale plantenmassa weer beginnen te groeien tot zij een niveau bereikt dat gelijk is aan de oorspronkelijke waarde. Het fenomeen zal zich herhalen.



Figuur 13.1: Oscillaties van de totale massa van planten $x(t)$ en van de totale massa van herbivoren $y(t)$ als functie van de tijd.

Lotka bestudeerde het model iets verder in een tweede artikel dat in 1920 werd gepubliceerd onder de titel *Ongedempte oscillaties afgeleid uit de wet van de massa-actie*. Hij legde uit waarom het systeem op een periodieke manier kon oscilleren. Dit volgt uit het feit dat het punt $(x(t), y(t))$ op een gesloten baan moet blijven in het vlak met x op de horizontale as en y op de verticale as; om precies te zijn in het kwadrant waar $x \geq 0$ en $y \geq 0$ liggen (figuur 13.2).

Als we vergelijking (13.1) delen door vergelijking (13.2), krijgen we na

¹De periode T hangt af van de beginvoorwaarden, maar Lotka beseftte dit feit pas in 1925.

enige herschikking

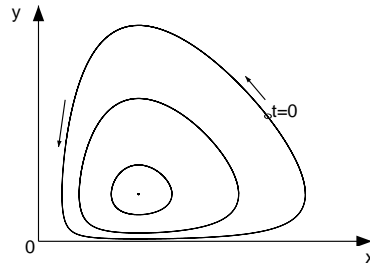
$$\left(-\frac{c}{x} + d\right) \frac{dx}{dt} = \left(\frac{a}{y} - b\right) \frac{dy}{dt}.$$

Integratie geeft

$$dx(t) - c \log x(t) = a \log y(t) - by(t) + K,$$

waarbij K een constante is die alleen afhangt van de begintoestand. Het punt $(x(t), y(t))$ blijft dus op de kromme $dx - c \log x = a \log y - by + K$, die toevallig een gesloten kromme is (figuur 13.2).

Figuur 13.2: Diagram met de totale massa van de planten $x(t)$ op de horizontale as en de totale massa van de herbivoren $y(t)$ op de verticale as. De drie gesloten krommen rond de stationaire toestand corresponderen met verschillende beginvoorwaarden.



De baan van $(x(t), y(t))$ draait rond de stationaire toestand $(c/d, a/b)$ tegen de wijzers van de klok in zoals gemakkelijk kan worden gezien door het teken van dx/dt en van dy/dt te bestuderen. In de buurt van de stationaire toestand vertoont het systeem kleine oscillaties met een periode gelijk aan $2\pi/\sqrt{ac}$.

Stel namelijk $x = \frac{c}{d} + x^*$ en $y = \frac{a}{b} + y^*$ waarbij $|x^*| \ll \frac{c}{d}$ en $|y^*| \ll \frac{a}{b}$. Dan is

$$\begin{aligned} \frac{dx^*}{dt} &= -by^* \left(\frac{c}{d} + x^*\right) \approx -\frac{bc}{d} y^*, \\ \frac{dy^*}{dt} &= dx^* \left(\frac{a}{b} + y^*\right) \approx \frac{ad}{b} x^*. \end{aligned}$$

Uit deze twee vergelijkingen verkrijgen we

$$\frac{d^2x^*}{dt^2} \approx -acx^* \quad \text{en} \quad \frac{d^2y^*}{dt^2} \approx -acy^*.$$

Deze vergelijkingen zijn dezelfde als voor de trillingen van de eenvoudige slinger in de natuurkunde. De periode is $2\pi/\sqrt{ac}$.

Raymond Pearl, die het eerste artikel uit 1920 had doorgegeven aan de *Proceedings of the National Academy of Sciences*, hielp Lotka een tweejarige beurs te krijgen van de Universiteit Johns Hopkins om een boek te schrijven met de titel *Elementen van de fysische biologie*. Het boek werd gepubliceerd in 1925. In het hoofdstuk waarin het werk uit 1920 werd samengevat, werd ook vermeld dat systemen van twee soorten, een gastheer en een parasietensoort of een prooi en een roofdierensoort, door hetzelfde model konden worden beschreven (13.1)–(13.2). Helaas trok Lotka's boek niet veel aandacht toen het werd gepubliceerd. De beroemde wiskundige Volterra herontdekte echter kort daarna onafhankelijk datzelfde model bij het bestuderen van een visserijprobleem.

Vito Volterra werd geboren in het Joodse getto van Ancona in 1860, kort voor de eenwording van Italië, toen de stad nog tot de Pauselijke Staten behoorde. Hij was een alleenstaand kind. Zijn vader, een lakenhandelaar, stierf toen Vito twee jaar oud was en liet het gezin zonder geld achter. Volterra, een goede leerling op de middelbare school, slaagde erin ondanks de armoede verder te studeren, eerst aan de universiteit van Florence en later aan de *Scuola Normale Superiore* in Pisa. In 1882 promoveerde hij in de natuurkunde en het jaar daarop werd hij professor in de mechanica aan de universiteit van Pisa. In 1892 trad hij toe tot de universiteit van Turijn, waarna hij in 1900 een leerstoel in de mathematische fysica ging bekleden aan de universiteit *La Sapienza* in Rome. In 1905 werd hij senator. Veel van de lezingen die hij in Rome of aan buitenlandse universiteiten gaf, werden in boekvorm gepubliceerd: *Drie lessen over enkele recente vorderingen in de mathematische fysica* (Universiteit van Clark, 1909), *Lessen over integrale en integro-differentiaalvergelijkingen* (Rome, 1910), *Lessen over lineaire functies* (Parijs, 1912), *De theorie van Permuteerbare Functies* (Princeton, 1912). Hij diende als officier in het Italiaanse leger tijdens de Eerste Wereldoorlog en leidde het bureau voor oorlogsuitvindingen. Na de oorlog nam hij actief deel aan de oprichting van de Italiaanse Mathematische Unie (1922) en van de Italiaanse Nationale Onderzoeksraad (1923), waarvan hij de eerste voorzitter werd. Hij werd ook voorzitter van de Internationale Commissie voor de wetenschappelijke studie van de Middellandse Zee (1923) en voorzitter van de Accademia dei Lincei

(1924). Een andere monografie, geschreven in samenwerking met J. Pérès, *Lessen over samenstelling en permuteerbare functies*, werd gepubliceerd in 1924.



Figuur 13.3: Volterra (1860–1940) ontving een doctoraat honoris causa van de Universiteit van Cambridge in 1900.

In 1925, op 65-jarige leeftijd, kreeg Volterra belangstelling voor een studie van de zoöloog Umberto D'Ancona, die later zijn schoonzoon zou worden, over het aandeel kraakbeenvissen (zoals haaien en roggen) die in de jaren 1905-1923 in drie havens van de Adriatische Zee bij de visserij werden aangevoerd: Triëst, Fiume² en Venetië. D'Ancona had geconstateerd dat het aandeel van deze vissen was toegenomen tijdens de Eerste Wereldoorlog, toen de visserij-inspanning was verminderd (tabel 13.1).

Tabel 13.1: Percentage kraakbeenvissen in de visserij in Triëst, Fiume en Venetië voor, tijdens en na de Eerste Wereldoorlog.

jaar	1910	1911	1912	1913	1914	1915	1916
Triëst	5,7	8,8	9,5	15,7	14,6	7,6	16,2
Fiume	-	-	-	-	11,9	21,4	22,1
Venetië	21,8	-	-	-	-	-	-
jaar	1917	1918	1919	1920	1921	1922	1923
Triëst	15,4	-	19,9	15,8	13,3	10,7	10,2
Fiume	21,2	36,4	27,3	16,0	15,9	14,8	10,7
Venetië	-	-	30,9	25,3	25,9	25,8	26,6

²Nu Rijeka in Kroatië.

Aangezien de meeste kraakbeenvissen roofdieren zijn van kleinere vis-
sen, leek het erop dat een vermindering van de visserij-inspanning de roof-
diersoorten bevoordeelde. Volterra, die niet op de hoogte was van het werk
van Lotka, verklaarde deze waarneming aan de hand van hetzelfde model

$$\frac{dx}{dt} = ax - bxy, \quad \frac{dy}{dt} = -cy + dxy,$$

waarbij $x(t)$ staat voor het aantal prooien en $y(t)$ voor het aantal roofdieren.
Hij merkte, net als Lotka, dat dit systeem periodiek kan oscilleren met een
periode T die afhangt van de begintoestand (x_0, y_0) . Hij merkte ook op dat

$$\frac{d}{dt} \log x = a - by, \quad \frac{d}{dt} \log y = -c + dx.$$

Integrerend over één periode T (zodat $x(0) = x(T)$ en $y(0) = y(T)$), verkreeg
hij

$$\frac{1}{T} \int_0^T y(t) dt = \frac{a}{b}, \quad \frac{1}{T} \int_0^T x(t) dt = \frac{c}{d}.$$

Het gemiddelde over één periode van zowel het aantal prooien als het aantal
predatoren is dus onafhankelijk van de beginvoorwaarden. Bovendien neemt,
als de visserij-inspanning afneemt, de groeisnelheid a van de prooien toe,
terwijl de uitsterfsnelheid c van de predatoren afneemt. Het gemiddelde van
 $x(t)$ daalt dus en het gemiddelde van $y(t)$ stijgt: het aandeel van de roofdieren
neemt toe. Dit is precies wat is waargenomen voor de visserijstatistieken van
de Adriatische Zee.

Volterra publiceerde zijn artikel eerst in het Italiaans in 1926. Een En-
gelse samenvatting verscheen een paar maanden later in *Nature*. Lotka infor-
meerde Volterra en andere wetenschappers over de prioriteit van zijn studie
van predator-prooi systemen. Maar zijn artikel uit 1920 en zijn boek uit 1925
zouden niet altijd genoemd worden. Lotka werkte toen al voor een verzeke-
ringsmaatschappij, zodat zijn werk zich concentreerde op menselijke demo-
grafie. Volterra bleef nog een decennium lang werken aan varianten van het
predator-prey systeem. Hij gaf een reeks lezingen in 1928-1929 aan het juist
opgerichte Instituut Henri Poincaré in Parijs. De aantekeningen van deze le-
zingen werden in 1931 gepubliceerd onder de titel *Lessen over de Wiskundige
Theorie van de Strijd om het Leven*. In 1935 publiceerde Volterra in samen-
werking met Umberto D'Ancona een ander boek over *Biologische associaties
vanuit wiskundig oogpunt*.

Hoewel het roofdier-prooi-model de visserijgegevens correct lijkt te ver-
klaren, was het debat over het realisme van vereenvoudigde modellen in de
ecologie nog maar net begonnen en is het nog steeds een onderwerp van

wetenschappelijke discussie. Tegenwoordig is het roofdier-prooi-model ook bekend als het Lotka-Volterra-model en is het een van de meest geciteerde modellen in de ecologie.

In 1931 weigerde Volterra trouw te zweren aan Mussolini. Hij verloor zijn hoogleraarschap aan de universiteit van Rome en werd uitgesloten van de Italiaanse wetenschappelijke academies, waarvan hij een van de beroemdste leden was. Van toen af aan verbleef hij hoofdzakelijk buiten Italië, reisde door Europa en gaf lezingen. Hij publiceerde met J. Pérès het eerste deel van een *Algemene theorie van de Functionalen* (1936) en samen met B. Hostinský een boek over *Oneindig kleine lineaire operaties* (1938). Hij overleed in Rome in 1940.

Verdere lectuur

1. Goodstein, J.R.: *The Volterra Chronicles, The Life and Times of an Extraordinary Mathematician 1860-1940*. American Mathematical Society (2007)
2. Guerraggio, A., Nastasi, P.: *Italian Mathematics between the Two World Wars*. Birkhäuser, Basel (2005)
3. Israel, G., Gasca, A.M.: *The Biology of Numbers – The Correspondence of Vito Volterra on Mathematical Biology*. Birkhäuser, Basel (2002)
4. Kingsland, S.E.: *Modeling Nature, Episodes in the History of Population Ecology*, 2nd edn. University of Chicago Press (1995)
5. Lotka, A.J.: Analytical note on certain rhythmic relations in organic systems. *Proc. Natl. Acad. Sci.* 6, 410–415 (1920) pnas.org
6. Lotka, A.J.: Undamped oscillations derived from the law of mass action. *J. Amer. Chem. Soc.* 42, 1595–1599 (1920) archive.org
7. Lotka, A.J.: *Elements of Physical Biology*. Williams & Wilkins, Baltimore (1925) archive.org
8. Volterra, V.: Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Accad. Lincei* 6, 31–113 (1926) → *Opere matematiche*, vol. 5, Accademia nazionale dei Lincei, Roma (1962) liberliber.it
9. Volterra, V.: Fluctuations in the abundance of a species considered mathematically. *Nature* 118, 558–560 (1926). → L.A. Real, J.H. Brown (eds.) *Foundations of Ecology*, 283–285. University of Chicago Press (1991)
10. Volterra, V.: *Leçons sur la théorie mathématique de la lutte pour la vie*. Gauthier-Villars, Paris (1931) gallica.bnf.fr
11. Volterra, V., D'Ancona, U.: *Les Associations biologiques au point de vue mathématique*. Hermann, Paris (1935)
12. Whittaker, E.T.: Vito Volterra 1860–1940. *Obit. Not. Fellows R. Soc.* 3, 690–729 (1941)

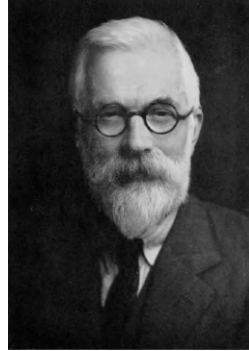
Hoofdstuk 14

Fisher en natuurlijke selectie (1922)

In 1922 publiceerde de Britse wiskundige bioloog Ronald Fisher een zeer invloedrijk artikel over populatiegenetica. Dit hoofdstuk behandelt slechts een deel van het artikel, dat zich richt op een variant van het Hardy-Weinberg model inclusief natuurlijke selectie. Fisher toonde aan dat als de heterozygoot bevoordeeld wordt, beide allelen naast elkaar kunnen bestaan. Als één van de twee homozygoten wordt bevoordeeld, dan verdwijnt het andere allel. De vraag is waarom sommige genen meerdere allelen kunnen hebben.

Ronald Aylmer Fisher werd in 1890 in Londen geboren als laatste van zes kinderen. Zijn vader was veilingmeester, maar verklaarde zich later failliet. Fisher studeerde tussen 1909 en 1913 wiskunde en natuurkunde aan het *Gonville and Caius College* van de universiteit van Cambridge. De genetica ontwikkelde zich in die tijd snel. Vanaf 1911 nam Fisher deel aan de bijeenkomsten van de *Eugenics Society* die door Galton was geïnitieerd. Hij begon zich te richten op statistische problemen die verband hielden met het werk van Galton en Mendel. Na zijn universitaire studies werkte hij een zomer op een boerderij in Canada en daarna voor de *Mercantile and General Investment Company* in Londen. Door zijn extreme bijziendheid kon hij niet deelnemen aan de Eerste Wereldoorlog, hoewel hij zich er vrijwillig voor had aangemeld. Hij bracht deze jaren door als leraar op middelbare scholen. In zijn vrije tijd zorgde hij voor een boerderij en zette hij zijn onderzoek voort. Hij verkreeg belangrijke nieuwe resultaten die correlatiecoëfficiënten in verband brachten met de Mendeliaanse genetica. In 1919 begon hij als statisticus te werken bij het Experimentenstation Rothamsted, dat zich richtte op landbouw.

In 1922 publiceerde Fisher een artikel getiteld *Over de dominantieverhouding*. Naast verschillende andere belangrijke nieuwe ideeën, beschouwde dit artikel een wiskundig model dat de wetten van Mendel combineerde met het idee van natuurlijke selectie waarop Darwin de nadruk legde voor de evolutietheorie. Fisher beschouwde dezelfde situatie als Hardy met twee allelen A en a en met de hypothese van willekeurige paring. Hij nam echter aan dat individuen met genotypes AA , Aa en aa een verschillende sterfte hebben voordat ze de volwassenheid bereiken, en zo natuurlijke selectie nabootsen. Stel p_n , $2q_n$ en r_n voor als de frequenties van de drie genotypes bij volwassen



Figuur 14.1:
Fisher (1890–1962)

individuen in generatie n , dan zijn er respectievelijk

$$(p_n + q_n)^2, \quad 2(p_n + q_n)(q_n + r_n) \quad \text{en} \quad (q_n + r_n)^2$$

pasgeborenen in generatie $n + 1$ met deze genotypes. Neem nu aan dat u , v en w de respectieve overlevingskansen van geboorte tot volwassenheid zijn. De frequenties van de genotypes bij volwassen individuen in generatie $n + 1$ zijn dan p_{n+1} , $2q_{n+1}$ en r_{n+1} met

$$p_{n+1} = \frac{u(p_n + q_n)^2}{d_n} \quad (14.1)$$

$$q_{n+1} = \frac{v(p_n + q_n)(q_n + r_n)}{d_n} \quad (14.2)$$

$$r_{n+1} = \frac{w(q_n + r_n)^2}{d_n}, \quad (14.3)$$

waar we voor het gemak aannemen dat

$$d_n = u(p_n + q_n)^2 + 2v(p_n + q_n)(q_n + r_n) + w(q_n + r_n)^2.$$

Met $p_n + 2q_n + r_n = 1$ in gedachten, zien we dat wanneer $u = v = w$ (d.w.z. wanneer er geen natuurlijke selectie is), het systeem (14.1)-(14.3) herleidt tot het systeem (11.1)-(11.3) van Hardy.

Stel $x_n = p_n + q_n$ is de frequentie van allel A bij volwassen individuen in generatie n . Dan is $q_n + r_n = 1 - x_n$ de frequentie van allel a . Als we (14.1) en (14.2) optellen, krijgen we

$$x_{n+1} = \frac{u x_n^2 + v x_n(1 - x_n)}{u x_n^2 + 2v x_n(1 - x_n) + w(1 - x_n)^2}.$$

Deze vergelijking kan worden herschreven in de vorm

$$x_{n+1} - x_n = x_n(1 - x_n) \frac{(v - w)(1 - x_n) + (u - v)x_n}{ux_n^2 + 2vx_n(1 - x_n) + w(1 - x_n)^2}. \quad (14.4)$$

Er zijn altijd minstens twee stabiele toestanden waarbij de frequentie x_n constant blijft tijdens de generaties: $x = 0$ (de populatie bestaat volledig uit homozygoot aa) en $x = 1$ (de populatie bestaat volledig uit homozygoot AA).

Met behulp van vergelijking (14.4) kan men aantonen dat als de homozygoot AA een betere overlevingskans heeft dan de twee andere genotypen ($u > v$ en $u > w$), het allel a geleidelijk uit de populatie zal verdwijnen. Dit geval zou in de natuur niet erg vaak mogen voorkomen als we weten dat beide allelen naast elkaar bestaan. Indien echter de heterozygoot Aa een selectief voordeel heeft ten opzichte van de homozygoten AA en aa ($v > u$ en $v > w$), dan kunnen de drie genotypes naast elkaar in de populatie bestaan. Dit is het meest voorkomende geval en het kan een verklaring zijn voor de „groei-kracht” van de hybriden die door de boeren wordt opgemerkt.

De stationaire toestand $x = 1$ is stabiel wanneer $u > v$ omdat

$$x_{n+1} - x_n \approx (1 - x_n)(u - v)/u$$

wanneer x_n dicht bij 1 ligt. De populatie neigt naar deze stationaire toestand. De stationaire toestand $x = 1$ is onstabiel als $u < v$, in welk geval er een derde stationaire toestand is

$$x^* = \frac{v - w}{2v - u - w}$$

met $0 < x^* < 1$. Bovendien kunnen we nagaan dat deze stabiel is. De stabiele toestand x^* komt overeen met een mengsel van de drie genotypen.

Door de wetten van Mendel eenvoudigweg te combineren met een hypothese van natuurlijke selectie (in dit geval verschillende overlevingskansen voor de drie genotypen), kunnen we de twee situaties van coëxistentie of verdwijning van genotypen verklaren. Na Fisher werd dit model ook ontwikkeld door J. B. S. Haldane (zie hoofdstuk 17) en door Sewall Wright (zie hoofdstuk 19).

Vooruitlopend op hoofdstuk 20, merk op dat als A volledig dominant is en de homozygoot aa benadeeld is ten opzichte van de twee andere genotypes, waarbij de getallen $u : v : w$ in een verhouding $1 : 1 : 1 - \varepsilon$ staan, dan wordt vergelijking (14.4)

$$x_{n+1} - x_n = \frac{\varepsilon x_n (1 - x_n)^2}{1 - \varepsilon (1 - x_n)^2} \approx \varepsilon x_n (1 - x_n)^2 \quad (14.5)$$

voor $\varepsilon \ll 1$. Als de overlevingskans van de heterozygoot Aa halverwege die van de twee homozygoten ligt, dan zijn de getallen $u : v : w$ in een verhouding $1 : 1 - \varepsilon/2 : 1 - \varepsilon$ en

$$x_{n+1} - x_n = \frac{\frac{\varepsilon}{2} x_n (1 - x_n)}{1 - \varepsilon (1 - x_n)} \approx \frac{\varepsilon}{2} x_n (1 - x_n) \quad (14.6)$$

als $\varepsilon \ll 1$.

Bij Rothamsted analyseerde Fisher langetermijngegevens over gewasopbrengsten en meteorologie. Hij leverde echter ook grote bijdragen aan de statistische methodologie. In 1925 publiceerde hij een boek getiteld *Statistische methoden voor onderzoekers*, dat veel succes had en vele malen herdrukt werd. In 1929 werd hij fellow van de *Royal Society*.

In 1930 publiceerde Fisher een boek getiteld *De genetische theorie van natuurlijke selectie*, een mijlpaal in de geschiedenis van de populatiegenetica. In 1933 werd hij hoogleraar eugenetica van *University College* in Londen, als opvolger van Karl Pearson aan het Galton Laboratorium. In 1943 werd hij hoogleraar genetica aan de Universiteit van Cambridge, ditmaal als opvolger van R. C. Punnett (zie hoofdstuk 11). Hij publiceerde ook verschillende boeken: *Het ontwerpen van experimenten* (1935), *De theorie van inteelt* (1949) en *Statistische methoden en wetenschappelijke gevolgtrekkingen* (1956). Hij werd in 1952 geridderd, vestigde zich na zijn pensionering in 1959 in Australië en overleed in 1962 in Adelaide. Op een ander deel van zijn werk komen we in hoofdstuk 20 terug.

Verdere lectuur

1. Fisher Box, J.: R.A. Fisher, *The Life of a Scientist*. Wiley (1978)
2. Fisher, R.A.: On the dominance ratio. *Proc. R. Soc. Edinb.* 42, 321–341 (1922) library.adelaide.edu.au
3. Fisher, R.A.: *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford (1930) archive.org
4. Yates, F., Mather, K.: *Ronald Aylmer Fisher 1890–1962. Biog. Mem. Fellows R. Soc.* 9, 91–120 (1963)

Hoofdstuk 15

Yule en evolutie (1924)

In 1924 bestudeerde de Britse statisticus Yule een evolutiemodel waarbij soorten nieuwe soorten kunnen voortbrengen door kleine mutaties en geslachten nieuwe geslachten kunnen voortbrengen door grote mutaties. Zijn doel was de verdeling van het aantal soorten binnen geslachten te verklaren, waarbij de meeste geslachten slechts één soort bevatten en een paar geslachten een groot aantal soorten. Het stochastische „geboorteproces” dat Yule in zijn model introduceerde is nog steeds een basisinstrument in de studie van fylogenetische bomen en vele andere gebieden.

George Udny Yule werd in 1871 in Schotland geboren. Zijn vader bekleedde een hoge functie in het Britse bestuur in India. Op 16-jarige leeftijd begon Yule te studeren aan het *University College* in Londen om ingenieur te worden. In 1892 veranderde hij van richting en deed hij een jaar onderzoek in Bonn onder supervisie van de natuurkundige Heinrich Hertz, die een paar jaar eerder het bestaan van elektromagnetische golven had aangetoond. Toen Yule terugkeerde naar Engeland, bood Karl Pearson hem een positie aan als universitair docent in de toegepaste wiskunde aan het *University College*. Yule begon zich, in navolging van Pearson, te richten op statistiek. In 1911 publiceerde hij *Een inleiding tot de theorie van de statistiek*, dat 14 maal herdrukt werd. Het jaar daarop verhuisde hij naar de universiteit van Cambridge. Zijn onderzoek had betrekking op theoretische aspecten van de statistiek, maar ook op toepassingen in de landbouw en de epidemiologie. In 1922 werd hij lid van de *Royal Society*.

In 1924 publiceerde Yule een artikel getiteld *Een wiskundige evolutietheorie gebaseerd op de conclusies van Dr. J. C. Willis*. Willis was een collega van de *Royal Society* die in 1922 een boek had gepubliceerd met de titel *Leeftijd en gebied, een studie in geografische distributie en oorsprong van soorten*. Hij had de verdeling van soorten over verschillende geslachten in de classificatie van planten en dieren bestudeerd. Uit de gegevens die hij had verzameld, bleek dat de meeste geslachten slechts één soort bevatten, dat steeds minder geslachten een groter aantal soorten bevatten en dat er nog een paar geslachten waren met een groot aantal soorten. In tabel 15.1 staan de

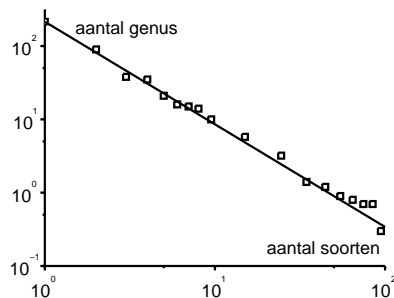


Figuur 15.1:
Yule (1871-1951)

gegevens betreffende slangen, hagedissen en twee keverfamilies (de *Chrysomelidae* en de *Cerambycinae*). De 1.580 toen bekende soorten hagedissen waren ingedeeld in 259 geslachten, waarvan 105 geslachten slechts één soort bevatten, 44 slechts twee soorten, 23 slechts drie soorten, enz. en twee geslachten met meer dan honderd soorten. Voor andere families van dieren en planten had de verdeling van de geslachten naar het aantal soorten dat zij bevatten een zeer vergelijkbare vorm.

Yule stelde voor dat Willis zou proberen zijn gegevens uit te zetten in een grafiek met logaritmische schalen. Dit leverde een opmerkelijk resultaat op (figuur 15.2): de logaritme van het aantal Q_n geslachten die n soorten bevatten, neemt min of meer lineair af met $\log(n)$. Met andere woorden, er zijn constanten $\alpha > 0$ en $\beta > 0$ zodanig dat $Q_n \approx \alpha n^{-\beta}$: de verdeling volgt een „machtswet”. In zijn artikel van 1924 zocht Yule naar een wiskundig evolutiemodel dat zo'n statistische verdeling kon verklaren.

Figuur 15.2: Het aantal geslachten als functie van het aantal soorten dat ze bevatten, met decimale logaritmische schalen. Gegevens voor de *Chrysomelidae*. Om de schommelingen bij een groot aantal soorten n af te vlakken, werden de geslachten geteld voor een interval van n -waarden, zoals in tabel 15.1. Het gemiddelde aantal geslachten voor één enkele waarde van n kan dus minder dan 1 zijn.



Tabel 15.1: Gegevens verzameld door Willis.

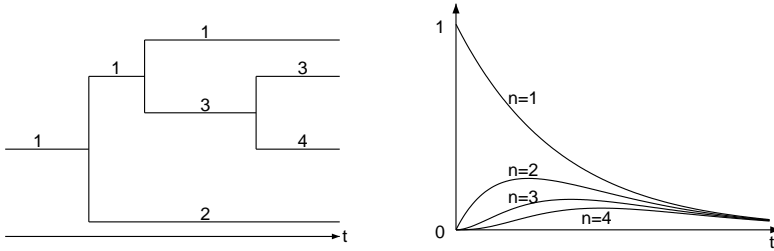
Aantal soorten	Aantal geslachten			
	<i>Chrysomelidae</i>	<i>Cerambycinae</i>	Slangen	Hagedissen
1	215	469	131	105
2	90	152	35	44
3	38	82	28	23
4	35	61	17	14
5	21	33	16	12
6	16	36	9	7
7	15	18	8	6
8	14	17	8	4
9	5	14	9	5
10	15	11	4	5
11-20	58	74	10	17
21-30	32	21	12	9
31-40	13	15	3	3
41-50	14	8	1	2
51-60	5	4	0	0
61-70	8	3	0	1
71-80	7	0	1	0
81-90	7	1	0	0
91-100	3	1	1	0
101-	16	4	0	2
totaal	627	1024	293	259

Daartoe bedacht hij eerst een stochastisch model¹ met continue tijd voor de groei van het aantal soorten binnen één geslacht (figuur 15.3a). Uitgaande van slechts één soort op het tijdstip $t = 0$, veronderstelde hij dat de kans voor een soort om door mutatie het leven te schenken aan een nieuwe soort van hetzelfde geslacht gedurende een „klein” tijdsinterval dt (op de tijdschaal van de evolutie) gelijk was aan $r dt$ met $r > 0$.

Zij $p_n(t)$ de waarschijnlijkheid dat er n soorten zijn op tijdstip t (n is een geheel getal maar t is een reëel getal). Om $p_n(t + dt)$ te berekenen, beschouwde Yule verschillende gevallen:

- als er op het tijdstip t $n - 1$ soorten zijn, heeft elke soort een kans $r dt$ om tussen t en $t + dt$ één nieuwe soort voort te brengen; in de limiet

¹McKendrick (zie hoofdstuk 16) was reeds begonnen met het bestuderen van dergelijke modellen in de populatiedynamica in een in 1914 gepubliceerd artikel.



Figuur 15.3: (a) Een simulatie van de evolutie van het aantal soorten binnen één geslacht. Soort 1 genereert soorten 2 en 3. Soort 3 genereert soort 4. (b) De kans $p_n(t)$ dat er n soorten van hetzelfde geslacht zijn op tijdstip t , voor $1 \leq n \leq 4$.

$dt \rightarrow 0$ zullen er n soorten zijn op het tijdstip $t + dt$ met een kans $(n - 1) r dt$;

- als er n soorten zijn op tijdstip t , zullen er $n + 1$ soorten zijn op tijdstip $t + dt$ met een kans van $n r dt$.

Dus $p_n(t)$ wordt gegeven door het volgende stelsel differentiaalvergelijkingen

$$\frac{dp_1}{dt} = -r p_1, \quad (15.1)$$

$$\frac{dp_n}{dt} = (n - 1) r p_{n-1} - n r p_n \quad (15.2)$$

voor alle $n \geq 2$. Uit de eerste vergelijking volgt $p_1(t) = e^{-rt}$ omdat $p_1(0) = 1$. Men kan aantonen dat de oplossing van de tweede vergelijking met $p_n(0) = 0$ gegeven wordt door

$$p_n(t) = e^{-rt} (1 - e^{-rt})^{n-1} \quad (15.3)$$

voor alle $n \geq 2$ (figuur 15.3b). Dus op een vast tijdstip t is de verdeling van de kansen $(p_n(t))_{n \geq 1}$ „geometrisch” en is de verhouding tussen twee opeenvolgende termen gelijk aan $1 - e^{-rt}$.

We merken immers eerst op dat vergelijking (15.2) equivalent is met

$$\frac{d}{dt} [p_n e^{nrt}] = (n - 1) r p_{n-1} e^{nrt}, \quad (15.4)$$

waaruit we achtereenvolgens $p_2(t)$, $p_3(t)$, ... kunnen berekenen. We krij-

gen $p_2(t) = e^{-rt}(1 - e^{-rt})$, dan $p_3(t) = e^{-rt}(1 - e^{-rt})^2$, wat formule (15.3) suggereert voor de algemene oplossing. Men kan tenslotte nagaan dat deze formule een oplossing is van vergelijking (15.4).

Yule leidde ook uit formule (15.3) af dat het verwachte aantal soorten exponentieel toeneemt met de tijd: $\sum_{n=1}^{+\infty} n p_n(t) = e^{rt}$.

We merken immers eerst op dat voor $|x| < 1$,

$$\sum_{n=1}^{+\infty} n x^{n-1} = \frac{d}{dx} \sum_{n=0}^{+\infty} x^n = \frac{d}{dx} \left(\frac{1}{1-x} \right) = \frac{1}{(1-x)^2}.$$

Dan

$$\sum_{n=1}^{+\infty} n p_n(t) = e^{-rt} \sum_{n=1}^{+\infty} n (1 - e^{-rt})^{n-1} = e^{rt}.$$

In het bijzonder, als T de verdubbelingstijd is, gedefinieerd door $e^{rT} = 2$, dan is de kansverdeling $(p_n(t))_{n \geq 1}$ van het aantal soorten op het tijdstip $t = T$ meetkundig met een verhouding $1/2$: $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16} \dots$. Op het tijdstip $t = kT$ is het meetkundig met een verhouding $1 - 1/2^k$ en $p_1(kT) = 1/2^k$.

Yule beschouwde vervolgens, parallel met de groei van het aantal soorten dat tot hetzelfde geslacht behoort, een soortgelijk proces als gevolg van grotere mutaties die leiden tot het ontstaan van nieuwe geslachten. Zij sdt de kans voor een bestaand geslacht om een nieuw geslacht te genereren gedurende een klein tijdsinterval dt . Als we ervan uitgaan dat er op het tijdstip $t = 0$ slechts één geslacht is, wordt het verwachte aantal geslachten op het tijdstip t gegeven door e^{st} . Het gemiddeld aantal geslachten dat per tijdseenheid op tijdstip t ontstaat is de afgeleide $s e^{st}$. In de limiet² $t \rightarrow +\infty$, is het gemiddelde aantal geslachten dat op tijdstip t tussen x en $x + dx$ tijdseenheden heeft bestaan dan $s e^{s(t-x)} dx$. De kans dat op tijdstip t een willekeurig gekozen geslacht tussen x en $x + dx$ tijdseenheden heeft bestaan is $s e^{s \cdot x} dx$.

Als een op tijdstip t willekeurig gekozen geslacht tussen x en $x + dx$ tijdseenheden heeft bestaan, is de kans dat dit geslacht n soorten bevat volgens formule (15.3) gelijk aan $e^{-rx}(1 - e^{-rx})^{n-1}$ voor alle $n \geq 1$. De kans q_n dat

²Yule beschouwde ook het geval waarin t niet erg groot kan worden verondersteld in vergelijking met de verdubbelingstijd van e^{rt} . De berekeningen zijn iets ingewikkelder maar de eindresultaten zijn niet erg verschillend.

een op tijdstip t willekeurig geslacht n soorten bevat, is dus

$$q_n = \int_0^{+\infty} s e^{-sx} e^{-rx} (1 - e^{-rx})^{n-1} dx.$$

Stel $u = r/s$. Een eenvoudige rekensom leert dat $q_1 = 1/(1+u)$ en dat

$$q_n = \frac{1}{1+u} \frac{u}{1+2u} \frac{2u}{1+3u} \cdots \frac{(n-1)u}{1+nu} \quad (15.5)$$

voor alle $n \geq 2$.

Er geldt immer dat $(1 - e^{-rx})^{n-1} = (1 - e^{-rx})^{n-2} (1 - e^{-rx})$. Dus

$$q_n = q_{n-1} - s \int_0^{+\infty} e^{-(r+s)x} (1 - e^{-rx})^{n-2} e^{-rx} dx.$$

Als we partieel integreren, krijgen we

$$q_n = q_{n-1} - \frac{r+s}{(n-1)r} q_n \quad \text{en} \quad q_n = \frac{(n-1)r/s}{1+n r/s} q_{n-1}.$$

Uit formule (15.5) blijkt dat de rij van kansen $(q_n)_{n \geq 1}$ een dalende rij is. Het maximum wordt dus bereikt voor $n = 1$: de meeste geslachten bevatten slechts één soort. Dit is precies wat de gegevens hadden aangetoond. Bovendien is de afname van q_n naar 0 als n naar oneindig gaat, relatief langzaam omdat $q_n/q_{n-1} \rightarrow 1$. Dit kan verklaren waarom sommige geslachten een groot aantal soorten bevatten. Om precies te zijn toonde Yule aan dat $\log q_n$ lineair afneemt met $\log(n)$.

Introduceer Euler's Gamma functie $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$. Dan $\Gamma(n+1) = n! = n \times (n-1) \times \cdots \times 2 \times 1$ als n een geheel getal is en $\Gamma(z+1) = z\Gamma(z)$. Dus (15.5) heeft de vorm

$$q_n = \frac{(n-1)!}{u(1+\frac{1}{u})(2+\frac{1}{u}) \cdots (n+\frac{1}{u})} = \frac{\Gamma(n)\Gamma(1+\frac{1}{u})}{u\Gamma(n+1+\frac{1}{u})}.$$

Maar Stirling's benadering geeft $\log \Gamma(n) \approx n \log n - n - \frac{1}{2} \log n + \text{constante}$. Op dezelfde manier, $\log \Gamma(n+1+1/u) \approx n \log n - n + (\frac{1}{u} + \frac{1}{2}) \log n + \text{constante}$. Ten slotte $\log q_n \approx -(1 + \frac{1}{u}) \log n + \text{constante}$.

Beschouw bijvoorbeeld het geval van de hagedissen. De parameter u kan geschat worden uit het aandeel $q_1 = 1/(1+u)$ van geslachten die slechts één soort bevatten. Volgens tabel 15.1 hebben we $q_1 = 105/259$ dus $u \approx 1,467$. We kunnen dan de theoretische kans q_n berekenen; vervolgens kunnen we het verwachte aantal Q_n van geslachten die n soorten bevatten, berekenen door q_n te vermenigvuldigen met het totale aantal soorten, dat 259 is (Tabel 15.2). Yule merkte op dat de overeenstemming tussen de waarnemingen en de berekeningen relatief goed³ is gezien de eenvoud van het model, dat bijvoorbeeld geen rekening houdt met de cataclysmen die soorten in de loop van miljoenen jaren evolutie hebben doorgemaakt.

Tabel 15.2: Vergelijking tussen gegevens en theorie in het geval van hagedissen (1.580 soorten ingedeeld in 259 geslachten).

Aantal soorten per geslacht	Waargenomen aantal geslachten	Berekend aantal geslachten
1	105	105,0
2	44	39,2
3	23	21,3
4	14	13,6
5	12	9,6
6	7	7,2
7	6	5,6
8	4	4,5
9	5	3,7
10	5	3,1
11-20	17	16,6
21-30	9	6,9
31-40	3	3,9
41-50	2	2,6
51-60	0	1,9
61-70	1	1,4
71-80	0	1,1
81-90	0	0,9
91-100	0	0,7
101-	2	10,1
totaal	259	259

³Voor het aantal geslachten met meer dan 100 soorten kreeg Yule een betere fit dan in tabel 15.2 door ervan uit te gaan dat t niet groot was in vergelijking met de verdubbelingstijd van e^{st} .

Na 1931 trok Yule zich geleidelijk terug van de Universiteit van Cambridge. Hij raakte geïnteresseerd in de statistische verdeling van de lengte van zinnen om schrijvers van boeken te identificeren. Hij paste dit in het bijzonder toe op het boek van John Graunt (zie hoofdstuk 2) dat mogelijk geïnspireerd is door William Petty. In 1944 publiceerde hij een boek over *De statistische studie van literaire woordenschat*. Hij overleed in 1951.

Tegenwoordig wordt het model van Yule nog steeds gebruikt voor de analyse van „fylogenetische bomen” (de genealogische bomen van soorten). Deze bomen, die lijken op die in figuur 15.3, zijn beter bekend dankzij de nieuwe gegevens uit de moleculaire biologie. Maar de toepassingen van het stochastisch proces gedefinieerd door vergelijkingen (15.1)–(15.2) zijn niet beperkt tot de evolutietheorie. Dit proces is een bouwsteen van vele modellen in de populatiedynamica, van het microscopisch niveau (om bijvoorbeeld kolonies van bacteriën te modelleren) tot het macroscopisch niveau (om het begin van een epidemie te modelleren). Het wordt wel „zuiver geboorteproses” of „Yule proces” genoemd. Een eenvoudige variant omvat een kans mdt op overlijden gedurende een klein tijdsinterval dt : de verwachte populatiegrootte op tijdstip t voor dit „geboorte- en sterfteproces” is dan $e^{(r-m)t}$. De kansverdeling (15.5) wordt ook wel de Yule-verdeling genoemd. Verdelingen met staarten die voldoen aan machtswetten hebben veel aandacht getrokken in verschillende gebieden van de wetenschap. De studie naar epidemieën in stochastische netwerken waar de graden verveeld zijn volgens een machtswet, is slechts één voorbeeld.

Verdere lectuur

1. Aldous, D.J.: Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. *Stat. Sci.* 16, 23–34 (2001) projecteuclid.org
2. Edwards, A.W.F.: George Udny Yule. In: Heyde, C.C., Seneta, E. (eds.) *Statisticians of the Centuries*, 292–294. Springer (2001)
3. McKendrick, A.G.: Studies on the theory of continuous probabilities with special reference to its bearing on natural phenomena of a progressive nature. *Proc. Lond. Math. Soc.* 13, 401–416 (1914)
4. Simon, H.A.: On a class of skew distribution functions. *Biometrika* 42, 425–440 (1955)
5. Willis, J.C.: *Age and Area*. Cambridge (1922) archive.org
6. Yates, F.: George Udny Yule. *Obit. Not. Fellows R. Soc.* 8, 308–323 (1952)
7. Yule, G.U.: A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, *Phil. Trans. Roy. Soc. Lond. B* 213, 21–87 (1925) gallica.bnf.fr

Hoofdstuk 16

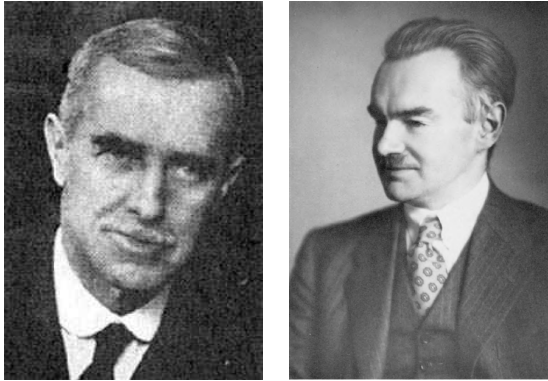
McKendrick en Kermack over epidemische modellering (1926–1927)

In 1926 bestudeerde McKendrick een stochastisch epidemiemodel en vond een methode om de kans te berekenen dat een epidemie een bepaalde uiteindelijke omvang bereikt. Hij ontdekte ook de partiële differentiaalvergelijking voor leeftijdsgestructureerde populaties in een continue-tijds kader. In 1927 bestudeerden Kermack en McKendrick een deterministisch epidemiemodel en verkregen zij een vergelijking voor de uiteindelijke omvang van de epidemie, waaruit blijkt dat er een bepaalde drempelwaarde is voor de bevolkingsdichtheid. Grote epidemieën kunnen wel boven, maar niet onder deze drempel optreden. Deze ideeën worden nog steeds veel gebruikt in de hedendaagse epidemiologie.

Anderson Gray McKendrick werd in 1876 in Edinburgh geboren als laatste van vijf kinderen. Hij studeerde geneeskunde aan de universiteit van Glasgow waar zijn vader professor in de fysiologie was. In 1900 trad hij in dienst bij de Indiase Medische Dienst. Voordat hij naar India ging, vergezelde hij Ronald Ross op een missie ter bestrijding van malaria in Sierra Leone. Daarna diende hij 18 maanden in het leger in Soedan. Bij zijn aankomst in India werd hij aangesteld als arts in een gevangenis in Bengalen waar hij dysenterie onder controle probeerde te krijgen. In 1905 trad hij in dienst bij het nieuwe Centraal Instituut voor Medisch Onderzoek in Kasauli (in het noorden van India). Hij werkte aan hondsdolheid maar studeerde ook wiskunde. In 1920 keerde hij na besmetting met een tropische ziekte terug naar Edinburgh en werd hij hoofd van het laboratorium van het *Royal College of Physicians*.

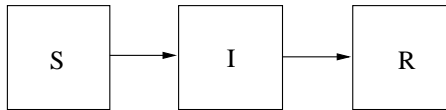
In 1926 publiceerde McKendrick een artikel over de *Toepassingen van de wiskunde op medische problemen*, dat verscheidene nieuwe ideeën bevatte. Hij introduceerde met name een continue-tijd wiskundig model voor epidemieën dat rekening hield met het stochastische aspect van infectie en herstel.

Beschouw een populatie van grootte N met aanvankelijk slechts één besmette persoon. Mensen kunnen achtereenvolgens drie toestanden doorlopen: de vatbare toestand S , de geïnfecteerde toestand I en de herstelde toestand R



Figuur 16.1: McKendrick (1876–1943) en Kermack (1898–1970)

(figuur 16.2)¹.



Figuur 16.2: Mogelijke staten: vatbaar (S), geïnfecteerd (I), hersteld (R).

Zij $p_{i,r}(t)$ de kans dat de bevolking op tijdstip t precies i personen in toestand I en r personen in toestand R bevat, waarbij i en r gehele getallen zijn zodanig dat $1 \leq i + r \leq N$. In dit geval wordt gezegd dat de bevolking zich in toestand (i, r) bevindt. Het aantal vatbare personen is dan $s = N - i - r$. In navolging van het werk van Ross over malaria (zie hoofdstuk 12), veronderstelde McKendrick dat gedurende een klein tijdsinterval dt de kans op een nieuwe infectie gelijk is aan $asidt$ (d.w.z. evenredig met zowel het aantal vatbare personen als het aantal geïnfecteerde personen). De kans op één nieuwe genezing is gelijk aan $bidt$. Zowel a als b zijn positieve parameters. Om $p_{i,r}(t + dt)$ te berekenen, moeten verschillende gevallen worden onderscheiden:

- de populatie bevindt zich in toestand $(i - 1, r)$ op tijdstip t en een nieuwe infectie brengt de populatie in toestand (i, r) tussen t en $t + dt$;

¹Het model van Daniel Bernoulli (zie hoofdstuk 4) omvatte de toestanden S en R maar niet I , omdat de duur van de besmetting veel korter is dan de gemiddelde levensverwachting.

de waarschijnlijkheid van deze gebeurtenis is $as(i-1)dt$ met $s = N - (i-1) - r$;

- de populatie bevindt zich in toestand (i, r) op tijdstip t en een nieuwe infectie brengt de populatie in toestand $(i+1, r)$ tussen t en $t+dt$; de waarschijnlijkheid van deze gebeurtenis is $asidt$ met $s = N - i - r$;
- de populatie bevindt zich in toestand $(i+1, r-1)$ op tijdstip t en een nieuw herstel brengt de populatie in toestand (i, r) tussen t en $t+dt$; de waarschijnlijkheid van deze gebeurtenis is $b(i+1)dt$;
- de populatie bevindt zich in toestand (i, r) op tijdstip t en één nieuw herstel brengt de populatie in toestand $(i-1, r+1)$ tussen t en $t+dt$; de waarschijnlijkheid van deze gebeurtenis is $bidt$.

Vandaar dat McKendrick de vergelijkingen

$$\begin{aligned} \frac{dp_{i,r}}{dt} = & a(N-i-r+1)(i-1)p_{i-1,r} - a(N-i-r)ip_{i,r} \\ & + b(i+1)p_{i+1,r-1} - b ip_{i,r} \end{aligned} \quad (16.1)$$

verkreeg voor $1 \leq i+r \leq N$. De eerste term aan het rechterlid ontbreekt als $i=0$, terwijl de derde term ontbreekt als $r=0$. De beginvoorwaarden zijn $p_{i,r}(0) = 0$ voor alle (i, r) behalve $p_{1,0}(0) = 1$.

Met dit model is McKendrick erin geslaagd de kans te berekenen dat de epidemie eindigt met n besmette personen, wat de limiet is van $p_{0,n}(t)$ wanneer $t \rightarrow +\infty$. Het is inderdaad niet nodig om systeem (16.1) op te lossen. Het volstaat op te merken dat zolang er i geïnfecteerden en r herstelden zijn, de kans op een nieuwe infectie gedurende een klein tijdsinterval dt $a(N-i-r)idt$ is en de kans op een nieuwe genezing $bidt$ is. De overgangskansen (zoals ze gewoonlijk genoemd worden in de theorie van Markovketens) van toestand (i, r) naar toestand $(i+1, r)$ of toestand $(i-1, r+1)$ zijn dus respectievelijk

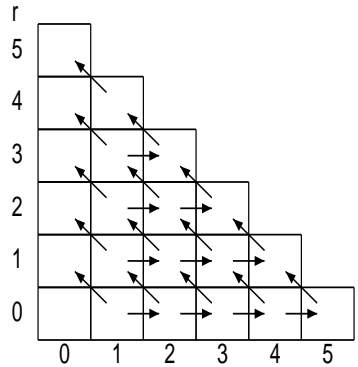
$$\mathcal{P}_{(i,r) \rightarrow (i+1,r)} = \frac{a(N-i-r)}{a(N-i-r)+b}, \quad \mathcal{P}_{(i,r) \rightarrow (i-1,r+1)} = \frac{b}{a(N-i-r)+b},$$

voor alle $i \geq 1$ (figuur 16.3).

Zij $q_{i,r}$ de kans dat de populatie tijdens de epidemie toestand (i, r) doorloopt. Aangezien $i=1$ en $r=0$ wanneer $t=0$, hebben we $q_{1,0} = 1$. De andere toestanden worden bereikt na een besmetting of na een herstel:

$$q_{i,r} = q_{i-1,r} \mathcal{P}_{(i-1,r) \rightarrow (i,r)} + q_{i+1,r-1} \mathcal{P}_{(i+1,r-1) \rightarrow (i,r)}.$$

Figuur 16.3: Diagram met de mogelijke toestanden van een populatie met $N = 5$ (i op de horizontale as, r op de verticale as) en de mogelijke overgangen ten gevolge van infectie (horizontale pijlen) of van herstel (andere pijlen).



De eerste term van het rechterlid ontbreekt als $i = 0$ of $i = 1$. De tweede term ontbreekt als $r = 0$. Uit deze formule kunnen we eerst $(q_{i,0})_{2 \leq i \leq N}$ berekenen, dan $(q_{i,1})_{0 \leq i \leq N-1}$, dan $(q_{i,2})_{0 \leq i \leq N-2}$ enz. De kans dat de epidemie uiteindelijk n mensen zal besmetten is $q_{0,n}$. In 1926 waren dergelijke berekeningen zeer omslachtig. Daarom beperkte McKendrick zich tot voorbeelden van zeer kleine populaties, bijvoorbeeld een gezin. Met $N = 5$ personen en $b/a = 2$ verkreeg hij Tabel 16.1. De grootste waarschijnlijkheden komen overeen met het geval waarin slechts één persoon in het gezin besmet is en met het geval waarin het hele gezin besmet is.

Tabel 16.1: Waarschijnlijkheid dat een epidemie in een gezin van vijf personen n besmet wanneer $b/a = 2$.

n	1	2	3	4	5
$q_{0,n}$	0,33	0,11	0,09	0,13	0,34

Hetzelfde artikel uit 1926 bevat ook een nieuwe formulering van demografische problemen wanneer tijd als een continue variabele wordt beschouwd. Voor dx oneindig klein, zij $P(x, t)dx$ de bevolking met een leeftijd tussen x en $x + dx$ op tijdstip t . Zij $m(x)$ het sterftcijfer op leeftijd x . Dan is

$$P(x+h, t+h) \approx P(x, t) - m(x)P(x, t)h$$

voor h oneindig klein. We introduceren de partiële afgeleiden van de functie $P(x, t)$:

$$\frac{\partial P}{\partial x}(x, t) = \lim_{h \rightarrow 0} \frac{P(x+h, t) - P(x, t)}{h}, \quad \frac{\partial P}{\partial t}(x, t) = \lim_{h \rightarrow 0} \frac{P(x, t+h) - P(x, t)}{h}.$$

Met behulp van de benadering

$$P(x+h, t+h) \approx P(x, t) + h \frac{\partial P}{\partial x}(x, t) + h \frac{\partial P}{\partial t}(x, t),$$

verkreeg McKendrick de volgende partiële differentiaalvergelijking:

$$\frac{\partial P}{\partial t}(x, t) + \frac{\partial P}{\partial x}(x, t) + m(x)P(x, t) = 0.$$

Een dergelijke vergelijking komt ook voor bij bevolkingsproblemen die gestructureerd worden door een continue variabele, zoals leeftijd in de demografie (zie hoofdstuk 25) of tijd sinds besmetting in de epidemiologie.

In 1921 werd William Ogilvy Kermack benoemd tot hoofd van de chemische afdeling van het *Royal College of Physicians Laboratory* in Edinburgh. Kermack werd in 1898 geboren in een klein stadje in Schotland. Hij studeerde aan de Universiteit van Aberdeen en begon onderzoek te doen op het gebied van de organische scheikunde in een industrieel laboratorium in Oxford. Ondanks het feit dat hij volledig blind werd na een explosie in zijn laboratorium in Edinburgh in 1924, zette hij zijn chemisch werk voort met de hulp van collega's en studenten. Kermack begon ook samen te werken met McKendrick aan de wiskundige modellering van epidemieën. Vanaf 1927 publiceerden zij samen een reeks *Bijdragen tot de mathematische theorie van epidemieën* waarin zij deterministische epidemiemodellen bestudeerden. Stel N is de bevolkingsgrootte met N groot genoeg. Veronderstel zoals in het artikel van 1926 dat mensen ofwel vatbaar, besmet of hersteld kunnen zijn. Indien de ziekte dodelijk is, is de derde toestand in feite de dood. Stel $S(t)$, $I(t)$ en $R(t)$ zijn het aantal mensen in elk van de drie toestanden. Het model is (in vereenvoudigde vorm) een stelsel van drie differentiaalvergelijkingen:

$$\frac{dS}{dt} = -aSI, \tag{16.2}$$

$$\frac{dI}{dt} = aSI - bI, \tag{16.3}$$

$$\frac{dR}{dt} = bI. \tag{16.4}$$

Het aantal nieuwe besmettingen per tijdseenheid is dus, zoals in het stochastische model van 1926, evenredig met zowel het aantal vatbare personen als het aantal besmette personen. Bij het begin van de epidemie, op tijdstip $t = 0$, is een bepaald aantal mensen besmet: $S(0) = N - I_0$, $I(0) = I_0$ en $R(0) = 0$, uitgaande van $0 < I_0 < N$.

Hoewel het systeem (16.2)-(16.4) geen gesloten oplossing heeft, kunnen de volgende conclusies getrokken worden:

- de totale bevolking $S(t) + I(t) + R(t)$ blijft constant en is gelijk aan N ;
- $S(t)$, $I(t)$ en $R(t)$ blijven nonnegatief (zoals het hoort aangezien het om populaties gaat);
- wanneer $t \rightarrow +\infty$, $S(t)$ afneemt tot een limiet $S_\infty > 0$, $I(t)$ naar 0 neigt en $R(t)$ toeneemt tot een limiet $R_\infty < N$;
- en bovendien de formule

$$-\log \frac{S_\infty}{S(0)} = \frac{a}{b}(N - S_\infty), \quad (16.5)$$

geeft impliciet S_∞ en dus ook de uiteindelijke epidemieomvang $R_\infty = N - S_\infty$.

Inderdaad, we zien eerst dat $\frac{d}{dt}(S + I + R) = 0$. Dus

$$S(t) + I(t) + R(t) = S(0) + I(0) + R(0) = N.$$

Vergelijkingen (16.2) en (16.3) kunnen worden herschreven als

$$\frac{d}{dt} \left[S(t) e^{a \int_0^t I(\tau) d\tau} \right] = 0, \quad \frac{d}{dt} \left[I(t) e^{bt - a \int_0^t S(\tau) d\tau} \right] = 0.$$

Hieruit volgt dat

$$S(t) = S(0) e^{-a \int_0^t I(\tau) d\tau} > 0, \quad I(t) = I(0) e^{a \int_0^t S(\tau) d\tau - bt} > 0.$$

Uit vergelijkingen (16.2) en (16.4) blijkt dan dat de functie $S(t)$ dalend is en dat de functie $R(t)$ stijgend is (in het bijzonder, $R(t) \geq 0$). Daar $S(t) \geq 0$ en $R(t) \leq N$, hebben de functies $S(t)$ en $R(t)$ wel degelijk limieten als $t \rightarrow +\infty$. Daar $I(t) = N - S(t) - R(t)$, heeft $I(t)$ ook een limiet als $t \rightarrow +\infty$, die alleen nul kan zijn zoals blijkt uit de integratie van (16.4).

Uit vergelijking (16.2) blijkt ook dat

$$-\frac{d}{dt}[\log S] = aI.$$

Als we integreren tussen $t = 0$ en $t = +\infty$, vinden we

$$\log S(0) - \log S_\infty = a \int_0^{+\infty} I(t) dt.$$

Vergelijking (16.3) kan worden herschreven als

$$\frac{dI}{dt} = -\frac{dS}{dt} - bI.$$

Als we integreren tussen $t = 0$ en $t = +\infty$, krijgen we

$$-I(0) = S(0) - S_\infty - b \int_0^{+\infty} I(t) dt.$$

Door de twee resultaten te combineren bekomen we formule (16.5), waaruit blijkt dat $S_\infty > 0$.

Wanneer het aanvankelijke aantal geïnfecteerden I_0 klein is in verhouding tot de bevolkingsomvang N , wat vaak het geval is bij het begin van een epidemie in een stad, kan formule (16.5) met $S_\infty = N - R_\infty$ worden herschreven als

$$-\log \left(1 - \frac{R_\infty}{N} \right) \approx \mathcal{R}_0 \frac{R_\infty}{N}, \quad (16.6)$$

waarbij per definitie

$$\mathcal{R}_0 = \frac{aN}{b}.$$

Vergelijking (16.6) heeft alleen een positieve oplossing als $\mathcal{R}_0 > 1$. Kermack en McKendrick komen dus tot de volgende conclusie: de epidemie infecteert slechts een niet-verwaarloosbaar deel van de bevolking als $\mathcal{R}_0 > 1$. Er is een drempel voor de bevolkingsdichtheid $N^* = b/a$ waaronder geen epidemieën kunnen voorkomen.

Wanneer de populatiegrootte N net boven deze drempel ligt ($N = N^* + \varepsilon$), treedt een epidemie van kleine amplitude op. Uit (16.6) volgt dat $R_\infty \approx 2\varepsilon$. Dus $S_\infty \approx N^* - \varepsilon$: de epidemie brengt de vatbare populatie even ver onder de drempel N^* als ze er aanvankelijk boven lag.

Als we de benadering $-\log(1-x) \approx x + \frac{x^2}{2}$ gebruiken, wordt vergelijking (16.6)

$$\frac{R_\infty}{N} + \frac{1}{2} \left(\frac{R_\infty}{N} \right)^2 \approx \mathcal{R}_0 \frac{R_\infty}{N}.$$

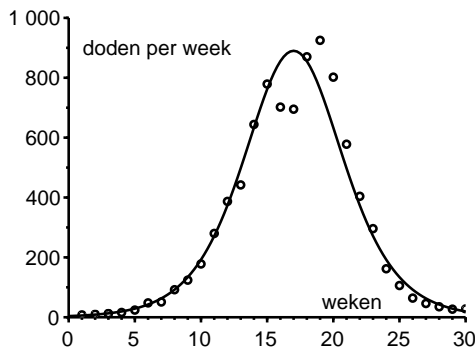
Dus

$$R_\infty \approx 2(\mathcal{R}_0 - 1)N = 2 \frac{\varepsilon}{N^*} (N^* + \varepsilon) \approx 2\varepsilon.$$

Zoals in het malariamodel van Ross (hoofdstuk 12) heeft de voorwaarde $\mathcal{R}_0 > 1$ een eenvoudige interpretatie. Aangezien aN het aantal personen is dat één besmette persoon per tijdseenheid besmet bij het begin van de epidemie en aangezien $1/b$ de gemiddelde besmettelijke periode is, is $\mathcal{R}_0 = aN/b$ het gemiddeld aantal secundaire gevallen ten gevolge van één besmette persoon bij het begin van de epidemie.

Voor dodelijke ziekten is $R(t)$ het cumulatieve aantal sterfgevallen sinds het begin van de epidemie en dR/dt het aantal sterfgevallen per tijdseenheid. Kermack en McKendrick hebben vastgesteld dat de grafiek van de functie dR/dt in hun wiskundig model de klokvorm heeft die men van een epidemie-curve verwacht (figuur 16.4).

Figuur 16.4: De curve dR/dt als functie van de tijd en de gegevens over het aantal sterfgevallen per week tijdens een pestepidemie in Bombay in 1905-1906.



Om dR/dt af te leiden deelden zij (16.2) door (16.4) om $dS/dR = -aS/b$ te verkrijgen. Dus

$$S(t) = S(0) \exp(-aR(t)/b).$$

Door dit in vergelijking (16.4) te vervangen en met behulp van $S(t) + I(t) + R(t) = N$, krijgen ze de vergelijking

$$\frac{dR}{dt} = b \left[N - R - S(0) \exp\left(-\frac{a}{b}R\right) \right], \quad (16.7)$$

die nog steeds niet expliciet kan worden opgelost. Als $R(t)$ echter gedurende de hele epidemie klein blijft, geeft de benadering $\exp(-u) \approx 1 - u + u^2/2$

$$\frac{dR}{dt} \approx b \left[N - R - S(0) + S(0) \frac{a}{b}R - S(0) \frac{a^2}{2b^2}R^2 \right]. \quad (16.8)$$

Dit is een zogenaamde Riccati vergelijking met twee constante oplossingen, een positieve R_+ en een negatieve R_- , gegeven door de wortels van de tweede-orde polynoom in R aan het rechterlid van (16.8). Zij $\tilde{R}(t)$ de exacte oplossing van (16.8) en stel $Q(t) = \tilde{R}(t) - R_+$. Dan voldoet $Q(t)$ aan een Bernoulli differentiaalvergelijking gelijk aan die van Daniel Bernoulli en Verhulst (zie (4.5) en (6.1)). Men kan dus formule (6.2) direct aanpassen om $Q(t)$ te krijgen. Een gemakkelijke maar vervelende berekening leert dat dQ/dt van de vorm

$$\frac{\alpha}{\cosh^2(\beta t - \gamma)}$$

is, waarbij α , β en γ constanten zijn die op een ingewikkelde manier afhangen van de parameters van het model. Aangezien $dR/dt \approx \tilde{R}/dt = dQ/dt$, konden Kermack en McKendrick (α, β, γ) kiezen om hun gegevens te passen. Moderne computers en software kunnen natuurlijk gemakkelijk de differentiaalvergelijking (16.7) numeriek oplossen zonder deze benaderingen te hoeven te doen.

De aldus verkregen curve voor dR/dt paste goed bij de gegevens over het aantal sterfgevallen per week tijdens de pestepidemie in Bombay tussen december 1905 en juli 1906 (figuur 16.4).

Kermack en McKendrick beschouwden ook het meer algemene model waarbij de besmettelijkheid $a(x)$ afhangt van de tijd x sinds de besmetting en waarbij het genezingspercentage $b(x)$ eveneens van x afhangt. De vergelijking die de uiteindelijke omvang van de epidemie geeft (wanneer het aanvankelijke aantal geïnfecteerde gevallen klein is) is nog steeds (16.6), maar

met

$$\mathcal{R}_0 = N \int_0^{+\infty} a(x) e^{-\int_0^x b(y) dy} dx. \quad (16.9)$$

De parameter \mathcal{R}_0 heeft dezelfde interpretatie als in het vorige geval: het is het gemiddeld aantal secundaire gevallen veroorzaakt door één besmette persoon aan het begin van de epidemie. Merk de overeenkomst op tussen (16.9) en de formule van Lotka (10.2) voor \mathcal{R}_0 in de demografie: leeftijd wordt vervangen door tijd sinds besmetting, overleving door de kans $e^{-\int_0^x b(y) dy}$ om nog besmet te zijn, vruchtbaarheid door de contactfrequentie $Na(x)$.

Kermack en McKendrick ontwikkelden in de jaren 1930 verschillende andere wiskundige modellen van epidemieën. Deze vormen nog steeds de bouwstenen voor de meeste van de complexere modellen die tegenwoordig in de epidemiologie worden gebruikt. De parameter \mathcal{R}_0 speelt nog steeds een centrale rol in de analyse van het model.

McKendrick ging in 1941 met pensioen en overleed in 1943. Tussen 1930 en 1933 was Kermack co-auteur van enkele artikelen over mathematische fysica met William McCrea en Edmund Whittaker, beiden van de afdeling wis-kunde van de Universiteit van Edinburgh. In de jaren 1930 en 1940 probeerde Kermack's team van scheikundigen nieuwe moleculen met antimalaria wer-king te synthetiseren, maar met beperkt succes. In 1938 schreef Kermack samen met Philip Eggleton een populair boek over elementaire biochemie, *Het materiaal waar we van gemaakt zijn*. In 1944 werd hij tot fellow van de *Royal Society* gekozen en in 1949 aanvaardde hij de leerstoel biochemie aan de Universiteit van Aberdeen. Later werd hij decaan van de Faculteit der Wetenschappen. Hij ging in 1968 met pensioen en overleed in 1970.

Verdere lectuur

1. Advisory Committee appointed by the Secretary of State for India, the Royal Society and the Lister Institute: Reports on plague investigations in India, XXII. *J. Hyg.* 7, 724–798 (1907) ncbi.nlm.nih.gov
2. Davidson, J.N., Yates, F., McCrea, W.H.: William Ogilvy Kermack 1898–1970. *Biog. Mem. Fellows R. Soc.* 17, 399–429 (1971)
3. Gani, J.: A.G. McKendrick. In: Heyde, C.C., Seneta, E. (eds.) *Statisticians of the Centuries*, 323–327. Springer, New York (2001)
4. Harvey, W.F.: A.G. McKendrick 1876–1943. *Edinb. Med. J.* 50, 500–506 (1943)
5. McKendrick, A.G.: Applications of mathematics to medical problems. *Proc. Edinb. Math. Soc.* 13, 98–130 (1926)
6. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* 115, 700–721 (1927) gallica.bnf.fr

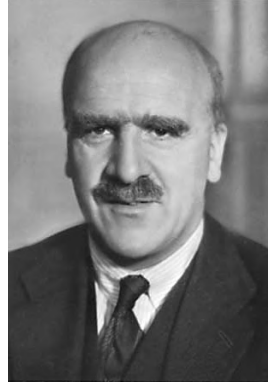
Hoofdstuk 17

Haldane en mutaties (1927)

In een ander deel van zijn artikel uit 1922 beschouwde Fisher het probleem van een mutant gen dat kan worden overgedragen op een willekeurig aantal nakomelingen met een gegeven waarschijnlijkheidsverdeling. Het probleem was formeel hetzelfde als dat van het uitsterven van familienamen, maar dan in een genetische context. Fisher toonde aan dat als de kansverdeling een Poisson-verdeling was en als het mutante gen geen selectief voordeel had, het mutante gen zeer langzaam uit de populatie kon verdwijnen. In 1927 zette de Britse bioloog Haldane de studie van dit model verder en toonde aan dat de waarschijnlijkheid dat een mutant gen zich zou handhaven tweemaal zijn selectief voordeel was. Hij gaf ook een meer rigoureuze behandeling van het uitstervingsprobleem.

John Burdon Sanderson Haldane werd in 1892 geboren in Oxford, waar zijn vader professor in de fysiologie was aan de universiteit. Haldane studeerde aan het *Eton College* en na 1911 aan het *New College* van de Universiteit van Oxford. Nadat hij zich in zijn eerste jaar op wiskunde had toegelegd, richtte hij zich op de menswetenschappen. Zijn studies werden onderbroken door de Eerste Wereldoorlog, waarin hij diende in Frankrijk en Irak. Nadat hij gewond was geraakt, werd hij als militair instructeur naar India gestuurd. In 1915 publiceerde hij een eerste artikel over genetische experimenten op muizen waarmee hij voor de oorlog was begonnen. In 1919 werd hij fellow van *New College*, waar hij fysiologie doceerde en net als zijn vader de ademhaling bestudeerde. In 1923 trad hij toe tot het biochemisch laboratorium van F. G. Hopkins¹ aan de universiteit van Cambridge, waar hij zich toelegde op de kinetica van enzymen. Hij publiceerde ook een science fiction roman, *Daedalus of Wetenschap en de Toekomst* (1923), en een essay getiteld *Calinicus, Een verdediging van chemische oorlogsvoering* (1925). Tussen 1924 en 1934 schreef hij een serie van tien artikelen onder de titel *Een wiskundige theorie van natuurlijke en kunstmatige selectie*.

¹Frederick Gowland Hopkins, die in 1929 de Nobelprijs voor fysiologie of geneeskunde kreeg voor zijn werk over vitamines.



Figuur 17.1:
Haldane (1892–1964)

In het vijfde artikel van de reeks, gepubliceerd in 1927, heroverwoog Haldane een ander genetisch model dat Fisher in 1922 had bestudeerd, een model gericht op mutaties. Fisher had de waarschijnlijkheid bestudeerd dat een mutant gen een populatie invadeert of verdwijnt. Dit probleem is formeel hetzelfde als dat van Bienaymé, Galton en Watson betreffende het uitsterven van familienamen. Maar Fisher verwees niet naar deze werken, hoewel hij misschien het artikel van Galton en Watson heeft gelezen dat in de appendix van Galton's boek *Natuurlijke erfenis* uit 1889 staat. Zoals in hoofdstuk 9, noem p_k de waarschijnlijkheid dat een gen wordt overgedragen op k nakomelingen in de eerste generatie ($k \geq 0$). Fisher beschouwde ook de genererende functie $f(x) = p_0 + p_1x + p_2x^2 + \dots + p_kx^k + \dots$, behalve dat hij geen bovengrens voor k vaststelde: de som kan een oneindig aantal termen bevatten. Hij beseftte dat, uitgaande van één individu met het mutante gen in generatie 0, de kans dat dit gen zich in k individuen bevindt, de coëfficiënt is van x^k in $f_1(x) = f(x)$ voor generatie 1, in $f_2(x) = f(f(x))$ voor generatie 2, in $f_3(x) = f(f(f(x)))$ voor generatie 3 enz. Op deze manier wordt duidelijk dat de volgende vergelijking geldt.

$$f_n(x) = f(f_{n-1}(x)) \quad (17.1)$$

Deze vergelijking is veel praktischer dan de door Watson afgeleide vergelijking $f_n(x) = f_{n-1}(f(x))$. In het bijzonder volgt uit (17.1) dat de uitsterfkans binnen n generaties $x_n = f_n(0)$ voldoet aan de iteratieformule $x_n = f(x_{n-1})$, zoals Bienaymé reeds had opgemerkt.

Als voorbeeld beschouwt Fisher het geval van een plant met een gemuteerd gen dat N zaden kan voortbrengen, waarbij elk zaadje een waarschijnlijkheid q heeft om te overleven en een nieuwe plant voort te brengen. De

kans p_k op k nakomelingen met het gemuteerde gen is binomiaal:

$$p_k = \binom{N}{k} q^k (1-q)^{N-k}$$

voor alle $0 \leq k \leq N$ en $p_k = 0$ voor $k > N$. De genererende functie is dan $f(x) = (1 - q + qx)^N$. Laat $\mathcal{R}_0 = Nq$ het gemiddelde aantal zaden zijn dat overleeft om een nieuwe plant voort te brengen. Als N groot is en q klein, dan

$$f(x) = \left(1 + \frac{\mathcal{R}_0}{N}(x-1)\right)^N \approx e^{\mathcal{R}_0(x-1)} = e^{-\mathcal{R}_0} \sum_{k=0}^{+\infty} \frac{(\mathcal{R}_0 x)^k}{k!}.$$

De kansverdeling (p_k) neigt naar $e^{-\mathcal{R}_0} (\mathcal{R}_0)^k / k!$, wat een Poisson-verdeling wordt genoemd. Fisher berekende dan de uitstervingskans binnen n generaties, gebruik makend van $x_0 = 0$, $x_n \approx e^{\mathcal{R}_0(x_{n-1}-1)}$ en de numerieke waarden $N = 80$ en $q = 1/80$. In dit geval is $\mathcal{R}_0 = Nq = 1$. Een omslachtige berekening leert dat $x_{100} \approx 0,98$: een mutant gen zonder selectief voordeel ($\mathcal{R}_0 = 1$) verdwijnt heel langzaam. Er is nog 2% kans dat het gen na 100 generaties nog in de populatie aanwezig is. In 1922 heeft Fisher de studie van dit model niet voortgezet.

Voortbordurend op het werk van Fisher, merkte Haldane in zijn artikel van 1927 voor het eerst op dat voor elke kansverdeling (p_k) zodanig dat $p_0 > 0$, de vergelijking $x = f(x)$ precies twee wortels heeft in het interval $(0, 1]$ wanneer het gemiddelde aantal nakomelingen dat drager is van het mutante gen \mathcal{R}_0 strikt groter is dan 1, d.w.z. wanneer het mutante gen een selectief voordeel heeft. Bovendien is de uitstervingskans x_∞ , die de limiet is van x_n als $n \rightarrow +\infty$, de kleinste van de twee wortels van $x = f(x)$: het gen heeft een positieve kans om zich in de populatie te vestigen. In tegenstelling tot Bienaymé en Cournot leverde Haldane een bewijs voor deze conclusie.

Inderdaad, $f'(x) \geq 0$ en $f''(x) \geq 0$ op het interval $[0, 1]$. Met andere woorden, de functie $f(x)$ is niet-degressief en convex. De veronderstellingen $f(0) = p_0 > 0$ en $f'(1) = \mathcal{R}_0 = p_1 + 2p_2 + 3p_3 + \dots > 1$ impliceren dat de vergelijking $f(x) = x$ precies twee oplossingen heeft in het interval $(0, 1]$: $x = 1$ en x^* zo dat $0 < x^* < 1$. Haldane verwees vervolgens naar een artikel van Gabriel Koenigs uit 1883, waarin werd aangetoond dat als $x_n = f(x_{n-1})$ en $x_n \rightarrow x_\infty$, dan $x_\infty = f(x_\infty)$ en $|f'(x_\infty)| \leq 1$. Als $f'(1) > 1$, is de enige mogelijkheid dat $x_\infty = x^*$.

Voor het geval van een Poisson-verdeling met $f(x) = e^{\mathcal{R}_0(x-1)}$ en \mathcal{R}_0 net iets groter dan 1, ligt de uitdovingskans x_∞ heel dicht bij 1. De vergelijking

$f(x_\infty) = x_\infty$ is dan equivalent met

$$\mathcal{R}_0(x_\infty - 1) = \log x_\infty \approx (x_\infty - 1) - \frac{(x_\infty - 1)^2}{2}.$$

Hieruit volgt dat

$$1 - x_\infty \approx 2(\mathcal{R}_0 - 1).$$

Haldane concludeerde dat de kans dat het gemuteerde gen niet zal uitsterven twee keer zo groot is als het selectieve voordeel $\mathcal{R}_0 - 1$. Zonder Haldane te citeren nam Fisher in zijn boek uit 1930 als voorbeeld het geval waarin $\mathcal{R}_0 = 1,01$, wat een kans van 2% oplevert dat het mutante gen niet zal uitsterven.

Haldane werd in 1932 lid van de *Royal Society*. Hij verliet Cambridge om professor in de genetica en later biometrie te worden aan het *University College* in Londen. Hij was toen vooral geïnteresseerd in menselijke genetica: schatting van mutatiepercentages, genetische kaarten van chromosomen enz. Naast zijn wetenschappelijke boeken (*Dierlijke biologie* in 1927 met Julian Huxley, *Enzymen* in 1930 en *De oorzaken van de evolutie* in 1932, *De biochemie van de genetica* in 1954), publiceerde hij een groot aantal artikelen over wetenschap in de pers (o.a. over de oorsprong van het leven) en enkele essays (*De ongelijkheid van de mens* in 1932, *De filosofie van een bioloog* in 1935, *De marxistische filosofie en de wetenschappen* in 1938, *Erfelijkheid en Politiek* in 1938 en *De vooruitgang van de wetenschap* in 1947). Na verscheidene bezoeken aan Spanje tijdens de burgeroorlog, probeerde hij zijn eigen land ervan te overtuigen schuilkelders te bouwen tegen luchtbombardementen. Tijdens de Tweede Wereldoorlog werkte hij aan ademhalingsproblemen in onderzeeërs. Lid van de communistische partij sinds 1942, nam hij in 1950 ontslag vanwege de officiële afwijzing van de Mendeliaanse genetica in de USSR door de invloed van Lysenko. In 1957 vestigde hij zich in India, waar hij zijn onderzoek voortzette, eerst aan het Indiase Statistische Instituut in Calcutta en later in Bhubaneswar. Nadat hij Indisch staatsburger was geworden, overleed hij in 1964.

Verdere lectuur

1. Clark, R.: *J.B.S., The Life and Work of J.B.S. Haldane*. London (1968)
2. Haldane, J.B.S.: A mathematical theory of natural and artificial selection, Part V, Selection and mutation. *Proc. Camb. Philos. Soc.* 23, 838–844 (1927)
3. Haldane, J.B.S.: *The Causes of Evolution*. Longmans (1932) archive.org
4. Pirie, N.W.: John Burdon Sanderson Haldane 1892-1964. *Biog. Mem. Fellows R. Soc.* 12, 218–249 (1966)

Hoofdstuk 18

Erlang en Steffensen over het uitstervingsprobleem (1929–1933)

In 1929 boog de Deense telefooningénieur Erlang zich opnieuw over het probleem van het uitsterven van familienamen. Zijn landgenoot, de statisticus Steffensen, werkte een volledige oplossing van het probleem uit. Hij toonde met name aan dat de verwachting van het aantal nakomelingen in elke generatie exponentieel toeneemt en sloeg daarmee de brug tussen stochastische en deterministische populatiemodellen.

Agner Krarup Erlang werd in 1878 geboren in Lønborg, Denemarken. Zijn vader was schoolmeester. Tussen 1896 en 1901 studeerde de jonge Erlang wiskunde, natuurkunde en scheikunde aan de Universiteit van Kopenhagen. Daarna gaf hij enkele jaren les op middelbare scholen, terwijl hij zich bleef interesseren voor wiskunde, vooral voor kansrekening. Hij ontmoette Jensen, hoofdingénieur bij de Telefoonmaatschappij van Kopenhagen en een amateur-wiskundige, die hem in 1908 overhaalde om bij het nieuwe onderzoekslaboratorium van de maatschappij te komen werken. Erlang begon artikelen te publiceren over de toepassingen van de waarschijnlijkheidsrekening op het beheer van telefoongesprekken. In 1917 ontdekte hij een formule voor wachttijden, die al snel door telefoonmaatschappijen over de hele wereld werd gebruikt. Zijn artikelen, die eerst in het Deens werden gepubliceerd, werden vervolgens in verschillende andere talen vertaald.

In 1929 raakte Erlang geïnteresseerd in hetzelfde probleem van uitsterven dat Bienaymé, Galton en Watson vóór hem hadden bestudeerd voor familienamen en dat Fisher en Haldane hadden bestudeerd voor mutante genen. Zoals zijn voorgangers was hij niet op de hoogte van alle werken die reeds waren gepubliceerd. Opnieuw p_k de waarschijnlijkheid noemend dat een individu k nakomelingen heeft, stelde hij vast dat de kans x_n op uitsterven binnen n generaties voldoet aan $x_n = p_0 + p_1 x_{n-1} + p_2 (x_{n-1})^2 + \dots = f(x_{n-1})$ met $x_0 = 0$. Hij merkte ook op dat de totale uitsterfkans x_∞ , die de limiet is van x_n als $n \rightarrow +\infty$, een oplossing is van de vergelijking $x_\infty = f(x_\infty)$. Hij beseftte dat $x = 1$ altijd een oplossing was en dat er een andere oplossing bestond tussen 0 en 1 wanneer het gemiddeld aantal nakomelingen $\mathcal{R}_0 = f'(1)$ groter is dan 1. Maar het lijkt erop dat hij niet kon achterhalen welke van deze twee oplos-



Figuur 18.1:
Erlang (1878–1929)

singen de juiste was. Net als Galton legde hij het probleem in 1929 voor in een Deens wiskundetijdschrift, *Matematisk Tidsskrift*:

„Vraag 15. Als de kans dat een individu k kinderen heeft p_k is, waarbij $p_0 + p_1 + p_2 + \dots = 1$, bepaal dan de kans dat zijn gezin uitsterft.”

Helaas overleed Erlang datzelfde jaar 1929 op 51-jarige leeftijd. Hij stierf overigens kinderloos¹.

Een hoogleraar actuariële wiskunde aan de Universiteit van Kopenhagen, Johan Frederik Steffensen, nam de vraag van Erlang over. Hij publiceerde in 1930 zijn oplossing in hetzelfde Deense tijdschrift: de uitsterfkans x_∞ is altijd de kleinste wortel van de vergelijking $x = f(x)$ in het gesloten interval $[0, 1]$, zoals Bienaymé en Haldane reeds hadden opgemerkt. Het bewijs van Steffensen is het bewijs dat in moderne leerboeken te vinden is.

We hebben immers gezien dat de uitsterfkans x_∞ een oplossing is van $x = f(x)$ in het gesloten interval $[0, 1]$. Laat x^* de kleinste oplossing zijn. Per definitie is $x^* \leq x_\infty$. Steffensen merkte als eerste op dat $x^* = f(x^*) \geq p_0 = x_1$. Neem bij inductie aan dat $x^* \leq x_n$. Dan geldt $x^* = f(x^*) \geq f(x_n) = x_{n+1}$ omdat de functie $f(x)$ stijgend is. Dus $x^* \geq x_n$ voor alle n . Als we de limiet nemen, dan is $x^* \geq x_n$. Dus $x_\infty = x^*$.

Steffensen gaf ook een meer formele verklaring waarom $x = 1$ de enige wortel is van $x = f(x)$ wanneer het gemiddeld aantal nakomelingen $\mathcal{R}_0 =$

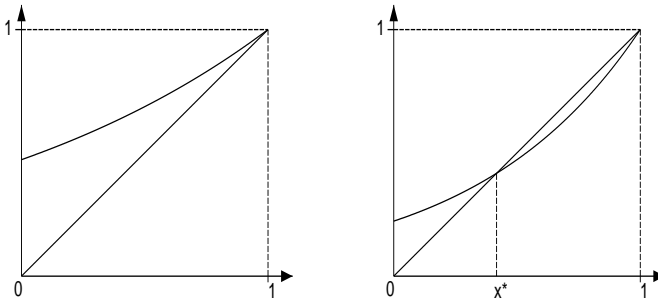
¹Ter nagedachtenis aan hem besloot het Internationaal Raadgevend Comité van de Telefoon in 1946 om „erlang” de meeteenheid van de intensiteit van het telefoonverkeer te noemen. „Erlang” is ook de naam die het bedrijf Ericsson aan een programmeertaal heeft gegeven.

$f'(1)$ kleiner of gelijk is aan 1 (figuur 18.2a) en waarom er slechts één andere wortel is dan $x = 1$ in het geval dat $\mathcal{R}_0 > 1$ (figuur 18.2b). Merk op dat $\mathcal{R}_0 = f'(1)$ de helling is van de functie $f(x)$ bij $x = 1$.

Hij merkte op dat voor elke wortel van $x = f(x)$, $1 - x = 1 - f(x) = 1 - p_0 - \sum_{k=1}^{+\infty} p_k x^k = \sum_{k=1}^{+\infty} p_k (1 - x^k)$. In de veronderstelling dat $x \neq 1$ en delend door $1 - x$, krijgen we

$$1 = p_1 + p_2(1+x) + p_3(1+x+x^2) + \dots \quad (18.1)$$

Als x toeneemt van 0 tot 1, neemt het rechterlid van vergelijking (18.1) toe van $1 - p_0$ tot $\mathcal{R}_0 = f'(1)$. Als $\mathcal{R}_0 < 1$, dan heeft vergelijking (18.1) geen oplossing. Als $\mathcal{R}_0 \geq 1$ en als we het triviale geval van $p_1 = 1$ uitsluiten, dan is het rechterlid van vergelijking (18.1) een strikt stijgende functie van x . Anders zou er geen $k \geq 2$ zijn zodanig dat $p_k \neq 0$ is, en zouden we dus $\mathcal{R}_0 = p_1 < 1$ hebben. Tot besluit, (18.1) heeft één en slechts één oplossing in het interval $[0, 1]$ als $\mathcal{R}_0 \geq 1$ is.



Figuur 18.2: Grafiek van de functies $y = x$ en $y = f(x)$ in het voorbeeld van hoofdstuk 17, $f(x) = e^{\mathcal{R}_0(x-1)}$, met $\mathcal{R}_0 = 0,75 < 1$ [links] of $\mathcal{R}_0 = 1,5 > 1$ [rechts].

Steffensen, die ook voorzitter was van het Deense Actuariële Genootschap en van het Deense Wiskundige Genootschap, werd in 1930 uitgenodigd aan de Universiteit van Londen. Zijn Britse collega W. P. Elderton vertelde hem over het werk van Galton en Watson. In 1933 publiceerde Steffensen een nieuw artikel in de annalen van het Instituut Henri Poincaré, waar hij in 1931 een conferentie had gegeven. Hij vatte de resultaten van zijn artikel in het Deens samen en vergeleek ze met die van Watson. Hij toonde ook aan dat de wiskundige verwachting van het aantal nakomelingen in generatie n gelijk is aan $(\mathcal{R}_0)^n$.

Laat $p_{k,n}$ de kans zijn dat er k nakomelingen zijn in generatie n , uitgaande van één individu in generatie 0. In zijn artikel uit 1930 had Steffensen net als zijn voorgangers opgemerkt dat voor de genererende functie van generatie n , $f_n(x) = \sum_{k=0}^{+\infty} p_{k,n} x^k$, geldt dat $f_1(x) = f(x)$ en

$$f_n(x) = f(f_{n-1}(x)). \quad (18.2)$$

Zij M_n de verwachting van het aantal nakomelingen in generatie n . Dan is $M_n = \sum_{k=1}^{+\infty} k p_{k,n} = f'_n(1)$. Als we formule (18.2) afleiden, krijgen we $f'_n(x) = f'(f_{n-1}(x)) \times f'_{n-1}(x)$. Dus $M_n = f'_n(1) = f'(f_{n-1}(1)) \times f'_{n-1}(1) = f'(1) \times M_{n-1} = \mathcal{R}_0 \times M_{n-1}$. Daar $M_1 = f'_1(1) = f'(1) = \mathcal{R}_0$, volgt daaruit dat $M_n = (\mathcal{R}_0)^n$ voor alle n .

Het verwachte aantal nakomelingen stijgt of daalt dus geometrisch naar gelang \mathcal{R}_0 groter of kleiner is dan 1. Het verwachte aantal nakomelingen gedraagt zich zoals in de deterministische modellen van bevolkingsgroei van Euler, Malthus enz. Echter, zelfs wanneer x_∞ groter is dan 1, is er een positieve kans dat de familie uitsterft. Deze mogelijkheid doet zich niet voor in deterministische modellen.

Het stochastisch proces dat door Steffensen en zijn voorgangers werd bestudeerd, is nog steeds het basiselement van vele meer realistische modellen van populatiedynamica. Wij zullen dit probleem nog een laatste keer aansnijden in hoofdstuk 20. Steffensen bleef hoogleraar aan de Universiteit van Kopenhagen tot 1943 en overleed in 1961.

Verdere lectuur

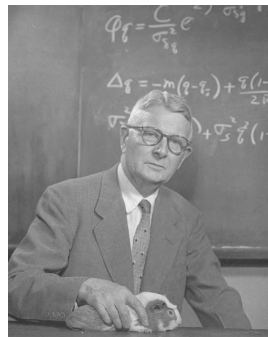
1. Brockmeyer, E., Halstrøm, H.L., Jensen, A.: The life and works of A.K. Erlang. *Trans. Dan. Acad. Techn. Sci.* 2 (1948)
2. Erlang, A.K.: Opgave Nr. 15. *Mat. Tidsskr. B*, 36 (1929) → Guttorp (1995)
3. Guttorp, P.: Three papers on the history of branching processes. *Int. Stat. Rev.* 63, 233–245 (1995) www.stat.washington.edu/research/reports/1992/tr242.pdf
4. Heyde, C.C.: Agner Krarup Erlang. In: Heyde, C.C., Seneta, E. (eds.) *Statisticians of the Centuries*, 328–330. Springer (2001)
5. Ogborn, M.E.: Johan Frederik Steffensen, 1873–1961. *J. R. Stat. Soc. Ser. A* 125, 672–673 (1962)
6. Steffensen, J.F.: Om Sandssynligheden for at Afkommet uddør. *Mat. Tidsskr. B*, 19–23 (1930) → Guttorp (1995)
7. Steffensen, J.F.: Deux problèmes du calcul des probabilités. *Ann. Inst. Henri Poincaré* 3, 319–344 (1933) archive.numdam.org

Hoofdstuk 19

Wright en willekeurige genetische drift (1931)

In 1931 ontwikkelde de Amerikaanse bioloog Sewall Wright het concept van een stochastisch model in de populatiegenetica, dat gebaseerd is op dezelfde veronderstellingen als in de wet van Hardy-Weinberg, behalve dat de populatie niet oneindig groot wordt verondersteld. De frequenties van de genotypen zijn niet langer constant. Een van de twee allelen zal in feite verdwijnen, maar misschien pas na een zeer lange tijd. De interpretatie van dit model bleef een twistpunt tussen Wright en Fisher, waarbij de laatste oordeelde dat natuurlijke selectie een belangrijkere rol speelt in de evolutie dan stochasticiteit.

Sewall Wright werd in 1889 in Massachusetts geboren. Hij studeerde af aan een klein *college* in Illinois waar zijn vader economie doceerde. Na een master in de biologie aan de Universiteit van Illinois in Urbana en een zomercursus aan het Cold Spring Harbor Laboratorium, promoveerde Wright aan de Harvard Universiteit op de overerving van vachtkleur bij de cavia. Tussen 1915 en 1925 bleef hij werken aan inteeltexperimenten met cavia's bij de Afdeling Dierhouderij van het Ministerie van Landbouw van de Verenigde Staten in Washington. Hij ontwikkelde de „methode van padcoëfficiënten” om deze experimenten te analyseren. Daarna ging hij werken bij de afdeling zoölogie van de Universiteit van Chicago.



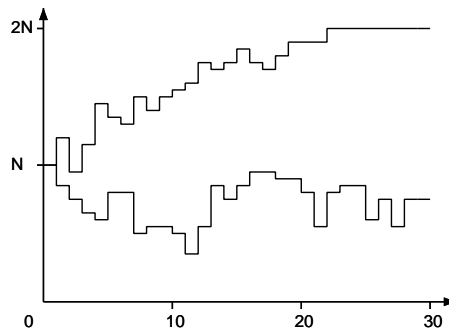
Figuur 19.1:
Wright (1889–1988)

Onder invloed van Fisher's artikel uit 1922 over populatiegenetica (zie

hoofdstuk 14) schreef Wright in 1925 een lang artikel getiteld *Evolutie in Mendeliaanse populaties*, dat uiteindelijk in 1931 werd gepubliceerd. Hij bestudeerde in het bijzonder een wiskundig model dat ook impliciet voorkwam in Fisher's boek uit 1930 *De Genetische Theorie van Natuurlijke Selectie*. Net als bij de wet van Hardy-Weinberg wordt in dit model het geval beschouwd waarin er slechts twee mogelijke allelen A en a voor één locus zijn, maar er wordt niet verondersteld dat de populatie oneindig groot is. De bedoeling is na te gaan of het weglaten van deze veronderstelling enige invloed heeft op de genetische samenstelling van de populatie. Laat N het totale aantal individuen zijn, waarvan wordt aangenomen dat het in alle generaties gelijk is. Elk individu heeft twee allelen, er zijn dus in totaal $2N$ allelen in de populatie in elke generatie. Het model gaat er ook van uit dat de paring willekeurig plaatsvindt. Als er in generatie n i allelen A en $2N - i$ allelen a zijn, dan zal een allel dat willekeurig wordt gekozen onder de individuen in generatie $n + 1$ A zijn met een kans $\frac{i}{2N}$ en a met een kans $1 - \frac{i}{2N}$. Het aantal A allelen in generatie $n + 1$ zal dus gelijk zijn aan j met een waarschijnlijkheid¹

$$p_{i,j} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j}, \quad (19.1)$$

waarbij $\binom{2N}{j}$ de binomiale coëfficiënt is die gelijk is aan $\frac{(2N)!}{j!(2N-j)!}$. Laat X_n het aantal A allelen in generatie n zijn: dit is een willekeurige variabele (fig. 19.2).



Figuur 19.2: Twee simulaties die de variaties tonen van het aantal X_n van A allelen gedurende 30 generaties als $N = 20$ en $X_0 = 10$.

Men kan aantonen dat de verwachting van X_{n+1} wetende dat $X_n = i$ gelijk

¹Wij danken deze formulering in de taal van Markov-ketens aan Malécot (1944).

is aan i : dit doet denken aan de wet van Hardy-Weinberg, waarbij de frequentie van allel A door de generaties heen constant blijft.

Inderdaad, beschouw de genererende functie

$$f(x) = \sum_{j=0}^{2N} p_{i,j} x^j = \left(1 - \frac{i}{2N} + \frac{ix}{2N}\right)^{2N},$$

De verwachting van X_{n+1} wetende dat $X_n = i$ is dan

$$\sum_{j=0}^{2N} j p_{i,j} = f'(1) = i. \quad (19.2)$$

In dit model is het echter mogelijk dat, uitgaande van een beginwaarde $X_0 = i$ met $0 < i < 2N$, de gebeurtenis $X_n = 0$ zich na een bepaald aantal generaties bij toeval voordoet. In dat geval zouden alle allelen van het type a zijn en zou X_n in alle toekomstige generaties gelijk blijven aan 0. Dezelfde fixatie zou gebeuren met allel A als $X_n = 2N$ na een bepaald aantal generaties. Samenvattend, wanneer de populatie oneindig groot wordt verondersteld zoals in het Hardy-Weinberg model, kunnen de twee allelen niet verdwijnen omdat hun frequenties constant blijven. Wanneer men rekening houdt met de eindige omvang van populaties, zoals in het Fisher-Wright model, fluctueren de frequenties van de twee allelen en kan (en zal) een van de allelen verdwijnen.

Uitgaande van $X_0 = i$ kan men gemakkelijk de kans Q_i berekenen dat de populatie gefixeerd wordt in toestand $X = 0$. Inderdaad, Q_i moet voldoen aan de „grensvoorwaarden”

$$Q_0 = 1, \quad Q_{2N} = 0. \quad (19.3)$$

Bovendien,

$$Q_i = \sum_{j=0}^{2N} p_{i,j} Q_j, \quad (19.4)$$

omdat $p_{i,j} Q_j$ de kans is om gefixeerd te worden in toestand $X = 0$ uitgaande van $X_0 = i$ en door $X_1 = j$ te gaan. Aangezien

$$\sum_{j=0}^{2N} p_{i,j} = 1,$$

zien we met behulp van (19.2) dat

$$Q_i = 1 - \frac{i}{2N}$$

de oplossing is van systeem (19.3)-(19.4). Vandaar dat de kans dat het systeem, uitgaande van i allelen van het type A in een populatie van grootte N , evolueert naar een populatie die alleen het allel a bevat, gelijk is aan $1 - \frac{i}{2N}$. Evenzo is de kans dat het systeem evolueert naar een populatie die alleen het allel A bevat gelijk aan $\frac{i}{2N}$.

Wright slaagde erin aan te tonen dat het aantal generaties dat verstrijkt vóór de fixatie in één van de twee extreme toestanden in de orde van $2N$ generaties ligt (fig. 19.3). Voor populaties van enkele miljoenen individuen zou deze tijd zo lang zijn dat de frequenties van de allelen als bijna constant kunnen worden beschouwd, zoals in de Hardy-Weinberg wet.

Stel dat er i_0 allelen van type A in de populatie van generatie 0 zijn. Laat $u_i^{(n)}$ de kans zijn dat er in generatie n i allelen van type A in de populatie zijn. Dan is

$$u_j^{(n+1)} = \sum_{i=0}^{2N} u_i^{(n)} p_{i,j}$$

voor alle $j = 0, \dots, 2N$. We hebben al gezien dat wanneer $n \rightarrow +\infty$,

$$u_0^{(n)} \rightarrow 1 - \frac{i_0}{2N}, \quad u_{2N}^{(n)} \rightarrow \frac{i_0}{2N}, \quad u_i^{(n)} \rightarrow 0$$

voor alle $0 < i < 2N$. Wright merkte op dat als $u_i^{(n)} = v$ voor alle $i = 1, \dots, 2N - 1$, dan

$$u_j^{(n+1)} = v \binom{2N}{j} \sum_{i=1}^{2N-1} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} \quad (19.5)$$

voor alle $1 < j < 2N$ omdat $p_{0,j} = p_{2N,j} = 0$. Wanneer N groot genoeg is,

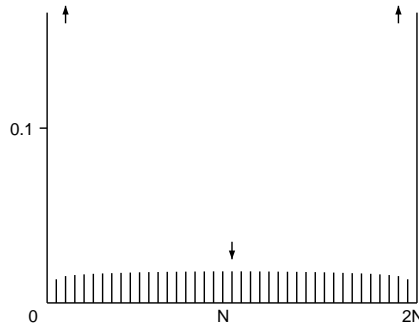
$$\begin{aligned} \frac{1}{2N} \sum_{i=1}^{2N-1} \left(\frac{i}{2N}\right)^j \left(1 - \frac{i}{2N}\right)^{2N-j} &\approx \int_0^1 x^j (1-x)^{2N-j} dx \\ &= \frac{j!(2N-j)!}{(2N+1)!}, \end{aligned} \quad (19.6)$$

de waarde van de integraal wordt verkregen door opeenvolgende partiële integraties. Door (19.5) en (19.6) te combineren komen we uiteindelijk voor $0 < j < 2N$ op

$$u_j^{(n+1)} \approx \frac{2N}{2N+1} v = \left(1 - \frac{1}{2N+1}\right) u_j^{(n)}.$$

De kansen $u_j^{(n)}$ voor alle $0 < j < 2N$ nemen dus af met een snelheid van ongeveer $1/2N$ per generatie. Deze snelheid is zeer traag als N groot is. Er is bijna geen afname als bijvoorbeeld N in de orde van grootte van miljoenen ligt.

Figuur 19.3: Kans dat er i allelen A in de populatie zijn ($i = 0, \dots, 2N$ op de horizontale as) na 30 generaties indien $N = 20$ en $X_0 = 10$.



In 1922 had Fisher reeds getracht deze fixatiesnelheid ($1/2N$) te schatten, maar hij had een factor 2 over het hoofd gezien. In elk geval waren de twee wetenschappers het niet eens over de typische grootte N van fokpopulaties. Voor de evolutietheorie suggereerde het werk van Wright dat willekeurige genetische drift in een kleine populatie een mechanisme voor het ontstaan van soorten zou kunnen zijn. Biologen die werkten aan de classificatie van soorten hadden inderdaad opgemerkt dat verschillen tussen soorten of ondersoorten vaak geen duidelijke verklaring hadden in termen van natuurlijke selectie. Dit idee werd in de jaren veertig en vijftig sterk bestreden door Fisher en zijn collega E.B. Ford, die beiden van mening waren dat willekeurige genetische drift te verwaarlozen was in vergelijking met natuurlijke selectie. Zij verwezen in het bijzonder naar hun studie van de fluctuaties van genfrequenties in een kleine geïsoleerde populatie motten (*Panaxia dominula*) in de buurt van Oxford, waar de drie genotypen voor een bepaald gen (gewone homozygoot,

heterozygoot en zeldzame homozygoot) op het oog konden worden onderscheiden. Een andere beroemde controversie over de respectievelijke invloed van natuurlijke selectie en willekeurige genetische drift spitste zich toe op slakken van het genus *Cepaea*. Realistischer evolutiemodellen combineren nu willekeurige genetische drift, selectie, mutatie, migratie, niet-willekeurige paring enz. De rol van willekeurige genetische drift werd later opnieuw benadrukt door de Japanse wetenschapper Motoo Kimura met zijn „neutrale theorie van moleculaire evolutie”. Een andere uitvloeisel was de ontwikkeling van de coalescentietheorie (geïntroduceerd door John Kingman in 1982), die de afstamming van genen terug in de tijd traceert tot het punt waar zij één gemeenschappelijke voorouder hebben.

Wright werd in 1934 lid van de Nationale Academie van Wetenschappen. Hij werkte vele jaren samen met Theodosius Dobzhansky aan de genetica van natuurlijke vliegenpopulaties (*Drosophila pseudoobscura*) in het *Death Valley* gebied. Hij ging in 1955 met pensioen aan de Universiteit van Chicago, maar bleef nog vijf jaar professor aan de Universiteit van Wisconsin-Madison. Tussen 1968 en 1978 publiceerde hij een vierdelige verhandeling die een samenvatting gaf van zijn werk over *Evolutie en de Genetica van Populaties*. Hij ontving de Balzan Prijs in 1984 en overleed in 1988 op 98 jarige leeftijd.

Verdere lectuur

1. Fisher, R.A.: *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford (1930) archive.org
2. Hill, W.G.: Sewall Wright, 21 December 1889–3 March 1988. *Biol. Mem. Fellows R. Soc.* 36, 568–579 (1990)
3. Kimura, M.: *The Neutral Theory of Molecular Evolution*. Cambridge University Press (1983)
4. Malécot, G.: Sur un problème de probabilités en chaîne que pose la génétique. *C. R. Acad. Sci. Paris* 219, 379–381 (1944)
5. Provine, W.B.: *Sewall Wright and Evolutionary Biology*. University of Chicago Press (1989)
6. Wright, S.: Evolution in Mendelian populations. *Genetics* 16, 97–159 (1931) www.esp.org
7. Wright, S.: *Evolution and the Genetics of Populations*, Vol. 2. University of Chicago Press (1969)

Hoofdstuk 20

De verspreiding van genen (1937)

In 1937 bestudeerden Ronald Fisher en drie Russische wiskundigen, Kolmogorov, Petrovskij en Piskunov, onafhankelijk van elkaar een partiële differentiaalvergelijking voor de geografische verspreiding van een voordelig gen. Zij toonden aan dat de genfrequentie zich gedroeg als een golf met een goed gedefinieerde snelheid, afhankelijk van het voordeel van het gen en een verspreidingscoëfficiënt. Hun werk was het uitgangspunt van de theorie van reactie-diffusievergelijkingen.

In 1937 werden twee artikelen gepubliceerd waarin een nieuwe benadering werd geïntroduceerd van de studie van ruimtelijke heterogeniteit in de populatiedynamiek. Fisher was de auteur van het eerste artikel, getiteld *De golf van opmars van genen met voordeel*, dat verscheen in *Annals of Eugenics*. Hij bestudeerde de ruimtelijke verspreiding van een gunstig gen in een populatie. Als vereenvoudiging beschouwde hij een ruimte met slechts één dimensie en noemde hij $u(x, t)$ het deel van de populatie dat zich op tijdstip t in het punt x bevindt en het gunstige gen bezit. Dus $0 \leq u(x, t) \leq 1$. Om natuurlijke selectie op te nemen, gebruikte hij vergelijking (14.6) met een continue tijdsvariabele $\frac{\partial u}{\partial t} = au(1-u)$, waarbij a een positieve parameter is. Voor een gegeven waarde van x herkennen we de logistische vergelijking van Verhulst (zie hoofdstuk 6) met een oplossing $u(x, t)$ die naar 1 neigt naarmate $t \rightarrow +\infty$. Bovendien nam Fisher aan dat de nakomelingen van een individu met het gunstige gen dat zich in het punt x bevinden, niet op hetzelfde punt blijven maar zich willekeurig verspreiden in de buurt van x . Naar analogie met de natuurkunde argumenteerde hij dat men een diffusie-term moet toevoegen aan de vergelijking voor $u(x, t)$, wat leidt tot de partiële differentiaalvergelijking

$$\frac{\partial u}{\partial t} = au(1-u) + D \frac{\partial^2 u}{\partial x^2}. \quad (20.1)$$

Wanneer de selectiecoëfficiënt a nul is, herleidt dit tot de diffusievergelijking die door Fourier in zijn warmteleer werd geïntroduceerd en later door Fick voor de diffusie van fysische deeltjes werd gebruikt. In 1904 begon Ronald Ross te denken over willekeurige verspreiding in de populatiedynamica. Hij vroeg zich toen af hoe de dichtheid van muggen afneemt naarmate de afstand

tot een broedplaats groter wordt. Het probleem was onder de aandacht gekomen van Karl Pearson en Lord Rayleigh. Tegen 1937 was de wetenschappelijke literatuur over diffusievergelijkingen aanzienlijk gegroeid, met name na Einsteins werk over Brownse beweging.

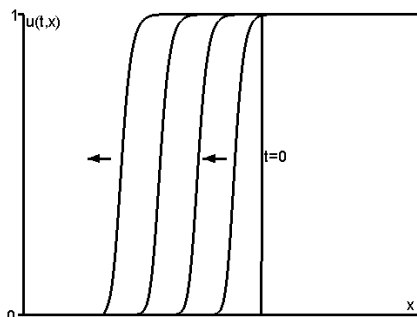
Fisher toonde aan dat er oplossingen bestaan van de vergelijking (20.1) van de vorm $u(x,t) = U(x + vt)$ die voldoen aan de drie voorwaarden

$$0 \leq u(x,t) \leq 1, \quad u(x,t) \xrightarrow{x \rightarrow -\infty} 0, \quad u(x,t) \xrightarrow{x \rightarrow +\infty} 1,$$

op voorwaarde dat $v \geq v^*$ waarbij

$$v^* = 2\sqrt{aD}.$$

Deze oplossingen verbinden de stationaire toestand $u = 1$ met het gunstige gen met de stationaire toestand $u = 0$ zonder dit gunstige gen. Zij stellen golven voor die zich met snelheid v voortplanten in de richting van afnemende waarden van x . Inderdaad, $u(x - vT, t + T) = u(x, t)$: het deel van de golf dat zich op tijdstip t in positie x bevond, beweegt zich naar positie $x - vT$ op tijdstip $t + T$.



Figuur 20.1: Voortplanting van rechts naar links van een gunstig gen met de snelheid v^* . De genfrequentie $u(t,x)$ op $t = 0$ is een stapfunctie.

Immers, door vast te stellen dat $z = x + vt$, merkte Fisher op dat als $u(x,t) = U(z)$, dan $\frac{\partial u}{\partial t} = vU'(z)$, $\frac{\partial u}{\partial x} = U'(z)$, $\frac{\partial^2 u}{\partial x^2} = U''(z)$. Als u een oplossing is van vergelijking (20.1), dan

$$vU'(z) = aU(z)(1 - U(z)) + DU''(z). \quad (20.2)$$

Wanneer u dicht bij 0 ligt, d.w.z. wanneer $z \rightarrow -\infty$, verwachtte Fisher

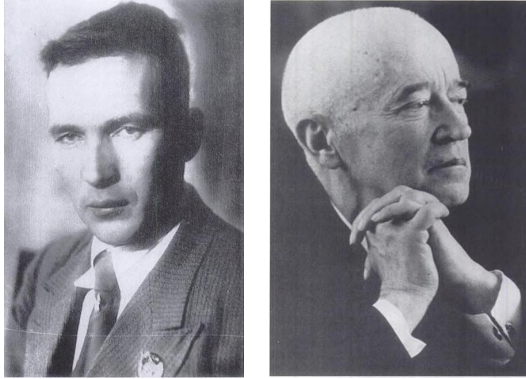
dat $U(z) \rightarrow 0$ en $U'(z) \rightarrow 0$. Als we k de limiet van $U'(z)/U(z)$ noemen wanneer $z \rightarrow -\infty$, dan weten we volgens de regel van L'Hôpital dat $U''(z)/U'(z) \rightarrow k$. Daarom, $U''(z)/U(z) = [U''(z)/U'(z)] \times [U'(z)/U(z)] \rightarrow k^2$. Het delen van vergelijking (20.2) door $U(z)$ en waarbij we z laten neigen naar $-\infty$, leid dan tot de tweede-ordevergelijking $Dk^2 - vk + a = 0$. Maar k moet een reëel getal zijn. De discriminant van deze vergelijking mag dus niet-negatief zijn: $v^2 - 4aD = 0$, of $v \geq 2\sqrt{aD} = v^*$. Vandaar dat $v \geq v^*$ een noodzakelijke voorwaarde is voor het bestaan van een golf die zich voortplant met de snelheid v . Het is ook een voldoende voorwaarde, zoals hieronder wordt uitgelegd.

Fisher merkte op dat alleen de golf die zich precies met de snelheid v^* voortplant, gekozen wordt voor een grote klasse van beginvoorwaarden, b.v. voor de stapfunctie: $u(x, 0) = 0$ voor $x < 0$, $u(x, 0) = 1$ voor $x \geq 0$. In figuur 20.1 is te zien hoe deze discontinue beginvoorwaarde geleidelijk overgaat in een vloeiende golf die zich voortplant in de richting van afnemende x met snelheid v^* .

In datzelfde jaar 1937, en onafhankelijk van het werk van Fisher, bestudeerden Andrej Nikolaevich Kolmogorov, Ivan Georgievich Petrovskij en Nikolaj Semënovich Piskunov hetzelfde probleem van de voortplanting van een dominant gen.

Kolmogorov werd in 1903 geboren in Tambov, Rusland. Tijdens zijn studie wiskunde aan de Staatsuniversiteit van Moskou verrichtte hij belangrijk werk op het gebied van goniometrische reeksen. Hij werd onderzoeker aan het Instituut voor Wiskunde en Mechanica in 1929 en universitair docent in 1931. Hij werkte aan stochastische processen en hun verband met differentiaalvergelijkingen en partiële differentiaalvergelijkingen. In 1933 publiceerde hij een verhandeling waarin hij de moderne grondslagen van de waarschijnlijkheidsrekening legde. Zijn onderzoeksinteresses omvatten topologie, approximatietheorie, Markovketens, Brownse beweging en ook toepassingen op biologische problemen. In 1935 publiceerde hij een artikel over genetica waarin hij de resultaten van Hardy, Fisher en Wright besprak. In 1936 publiceerde hij een artikel over een veralgemening van het Lotka-Volterra systeem.

Petrovskij werd in 1901 geboren in Sevsk. Hij studeerde ook wiskunde aan de Staatsuniversiteit van Moskou, waar hij in 1933 hoogleraar werd. Hij werkte vooral aan de theorie van partiële differentiaalvergelijkingen en aan de topologie van reële algebraïsche krommen, maar schreef ook enkele artikelen over gewone differentiaalvergelijkingen en over de theorie van de waarschijnlijkheid. Piskunov, geboren in 1908, was een andere oud-student wiskunde aan de Staatsuniversiteit van Moskou.



Figuur 20.2: Kolmogorov (1903–1987) en Petrovskij (1901–1973)

In de jaren dertig had Kolmogorov contacten met A. S. Serebrovskij, een pionier op het gebied van de populatiegenetica in Moskou. Het werd toen steeds gevaarlijker om de Mendeliaanse genetica in de USSR te verdedigen door de opkomst van Lysenko, een landbouwkundige die erin geslaagd was Stalin ervan te overtuigen dat de Mendeliaanse genetica slechts een „bourgeois pseudowetenschap” was. Het Zevende Internationale Congres van de Genetica, oorspronkelijk gepland voor 1937 in Moskou, werd afgelast. Veel Sovjetgenetici werden geëxecuteerd of naar werkkampen gestuurd.

In hun artikel uit 1937, getiteld *Een studie van de diffusievergelijking met toename van de hoeveelheid stof en de toepassing ervan op een biologisch probleem*, dat werd gepubliceerd in het Bulletin van de Moskouse Staatsuniversiteit, gebruikten Kolmogorov, Petrovskij en Piskunov niettemin een wiskundig model dat was gebaseerd op de Mendeliaanse genetica. Hun model was een partiële differentiaalvergelijking van de vorm

$$\frac{\partial u}{\partial t} = f(u) + D \frac{\partial^2 u}{\partial x^2} \quad (20.3)$$

waarbij $u(x, t)$ weer de frequentie is van het gunstige gen op punt x en tijdstip t . De functie $f(u)$ wordt verondersteld aan verschillende voorwaarden te voldoen: $f(0) = f(1) = 0$, $f(u) > 0$ als $0 < u < 1$, $f'(0) > 0$ en $f'(u) < f'(0)$ als $0 < u \leq 1$. De auteurs toonden een resultaat aan dat analoog is aan dat van Fisher, maar met een meer rigoureuus bewijs: als de beginvoorwaarde zodanig is dat $0 \leq u(x, 0) \leq 1$, $u(x, 0) = 0$ voor alle $x < x_1$ en $u(x, 0) = 1$ voor alle $x > x_2 \geq x_1$, dan plant het gen zich voort met de snelheid $v^* = 2\sqrt{f'(0)D}$.

Het zoeken naar een oplossing $u(x,t) = U(z)$ waarbij $z = x + vt$ leidt tot de voor de hand liggende veralgemening van vergelijking (20.2): $vU'(z) = f(U(z)) + DU''(z)$. Deze tweede-orde differentiaalvergelijking kan worden herschreven als een stelsel van eerste-orde differentiaalvergelijkingen

$$\frac{dU}{dz} = p, \quad \frac{dp}{dz} = \frac{vp - f(U)}{D}. \quad (20.4)$$

Herinner u dat $U(z)$ zo moet zijn dat $U(z) \rightarrow 0$ gaat als $z \rightarrow -\infty$ gaat en $U(z) \rightarrow 1$ als $z \rightarrow +\infty$ gaat. Nabij de stationaire toestand ($U = 0, p = 0$) van systeem (20.4) hebben we $f(U) \approx f'(0)U$. Dus (20.4) kan benaderd worden door het lineaire systeem

$$\frac{dU}{dz} = p, \quad \frac{dp}{dz} = \frac{vp - f'(0)U}{D}. \quad (20.5)$$

Het zoeken naar exponentiële oplossingen van de vorm $U(z) = U_0 e^{kz}$ en $p(z) = p_0 e^{kz}$ levert de karakteristieke vergelijking $Dk^2 - vk + f'(0) = 0$ op, zoals in Fisher's artikel. Ook hier moet k reëel zijn (anders zou u gaan oscilleren en negatieve waarden aannemen). Dus $v \geq 2\sqrt{f'(0)D} = v^*$. De twee wortels voor k zijn dan reëel en positief. Indien $v > v^*$, zijn de twee wortels verschillend en is de stationaire toestand ($U = 0, p = 0$) een onstabiel knooppunt. Indien $v = v^*$, zijn de twee wortels identiek en is ($U = 0, p = 0$) een onstabiel ontaard knooppunt zoals weergegeven in figuur 20.3. Op dezelfde manier leidt systeem (20.4) nabij de stationaire toestand ($U = 1, p = 0$) tot het lineaire systeem

$$\frac{d(U-1)}{dz} = p, \quad \frac{dp}{dz} = \frac{vp - f'(1)(U-1)}{D}$$

en tot de karakteristieke vergelijking $Dk^2 - vk + f'(1) = 0$. De discriminant is $v^2 - 4Df'(1) \geq 0$ omdat $f'(1) \leq 0$. Als $f'(1) < 0$, zijn er twee reële wortels met een tegengesteld teken en is ($U = 1, p = 0$) een zadelpunt. Als $f'(1) = 0$, is één wortel nul en de andere positief (zie fig. 20.3). Een gedetailleerde analyse toont aan dat er in feite voor alle $v \geq 2\sqrt{f'(0)D}$ een unieke integraalkromme bestaat die de twee stationaire toestanden ($U = 0, p = 0$) en ($U = 1, p = 0$) verbindt, zoals in het speciale geval van figuur 20.3.

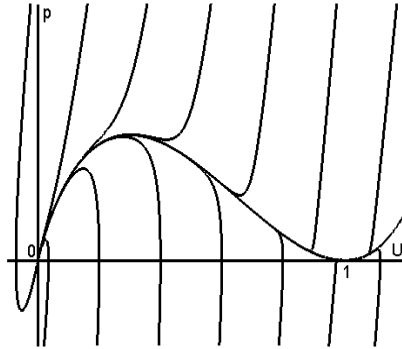
Kolmogorov, Petrovskij en Piskunov toonden vervolgens rigourens aan dat

- de partiële differentiaalvergelijking (20.3) een unieke oplossing $u(x,t)$ heeft die voldoet aan de beginvoorwaarde,
- Deze oplossing is zodanig dat $0 < u(x,t) \leq 1$ voor alle x en $t > 0$,
- en zodanig dat $u(x,t)$ een stijgende functie van x blijft als dat ook het geval is op $t = 0$ en
- zodanig dat $u(x,t)$ convergeert naar een golfprofiel dat zich voortplant met de snelheid v^* .

De bewijzen zijn te lang om hier samen te vatten.

Merk op dat de door Fisher gebruikte functie $f(u) = au(1-u)$ aan al deze voorwaarden voldoet met $f'(0) = a$. Geïnspireerd door vergelijking (14.5) beschouwden Kolmogorov, Petrovskij en Piskunov de functie $f(u) = au(1-u)^2$, die aan dezelfde voorwaarden voldoet en dezelfde voortplantingssnelheid geeft.

Figuur 20.3: Diagram (U, p) met enkele integraalkrommen van het systeem (petrovskiisysteem) en in het bijzonder de unieke kromme die $(U = 1, p = 0)$ verbindt met $(U = 0, p = 0)$, en die de vorm geeft van de voortplantingsgolf. Hierbij is $f(u) = au(1-u)^2$, $a = 1$, $D = 1$ en $v = v^* = 2$.



De artikelen van Fisher en van Kolmogorov, Petrovskij en Piskunov vormden het uitgangspunt voor de constructie van veel wiskundige modellen met geografische verspreiding in de genetica, ecologie en epidemiologie. Deze modellen staan bekend als „reactie-diffusie systemen”.

Kolmogorov bestudeerde vanaf 1938 ook het probleem van het uitsterven van familienamen van Bienaimé, Galton, Watson, Fisher, Haldane, Erlang en Steffensen: hij noemde het stochastische proces dat al deze werken gemeen

hebben het „takkenproces”. In 1939 werd hij lid van de USSR Academie van Wetenschappen. Later leverde hij belangrijke bijdragen aan het probleem van turbulentie in de vloeistofmechanica (1941), aan de theorie van dynamische systemen in verband met de hemelmechanica (1953) en aan de informatietheorie (vanaf 1956). Hij heeft ook bijgedragen aan het schrijven van een encyclopedie en van leerboeken voor middelbare scholen en universiteiten, heeft geholpen met het oprichten van een experimentele middelbare school en was redacteur van een populairwetenschappelijk tijdschrift. Hij ontving vele internationale prijzen (waaronder de Balzanprijs in 1963 en de Wolfprijs in 1980) en overleed in Moskou in 1987.

Petrovskij werd decaan van de Mechanica en Wiskunde Faculteit van de Moskouse Staatsuniversiteit in 1940. Hij was rector van de universiteit van 1951 tot aan zijn dood in 1973. Hij was een volwaardig lid van de USSR Academie van Wetenschappen vanaf 1946 en voorzitter van het Internationale Congres van Wiskundigen dat in 1966 in Moskou werd gehouden. Hij schreef ook leerboeken over gewone differentiaalvergelijkingen, partiële differentiaalvergelijkingen en integraalvergelijkingen. Piskunov werd professor aan een militaire academie. Zijn leerboek over differentiaal- en integraalrekeningen werd door vele technische universiteiten gebruikt. Hij overleed in 1977.

Verdere lectuur

1. Fisher, R.A.: The wave of advance of advantageous genes. *Ann. Eugen.* 7, 355–369 (1937) digital.library.adelaide.edu.au
2. Kolmogorov, A.N., Petrovskii, I.G., Piskunov, N.S.: Étude de l'équation de la diffusion avec croissance de la quantité de matière et son application à un problème biologique. *Bull. Univ. État Moscou Math. Mec.* 1, 1–26 (1937) → V.M. Tikhomirov (ed.) *Selected Works of A. N. Kolmogorov*, vol. 1, 242–270. Kluwer (1991).
3. Oleinik, O.A.: I.G. Petrowsky and modern mathematics. In: *I. G. Petrowsky Selected Works*, Part I, 4–30. Gordon and Breach, Amsterdam (1996)
4. Pearson, K.: *Mathematical Contributions to the Theory of Evolution, XV, A Mathematical Theory of Random Migration*. Dulau, London (1906) archive.org
5. Rosenfeld, B.A.: Reminiscences of Soviet Mathematicians. In: Zdravkovska, S., Duren, P.L. (eds.) *Golden Years of Moscow Mathematics*, 2nd edn., 75–100. Am. Math. Soc. (2007)
6. Shiryaev, A.N.: *Selected Works of A. N. Kolmogorov*, vol. 2. Kluwer (1992)
7. Shiryaev, A.N.: Andrei Nikolaevich Kolmogorov (April 25, 1903 to October 20, 1987). In: *Kolmogorov in Perspective*, 1–88. Am. Math. Soc. (2000)

Hoofdstuk 21

De Leslie-matrix (1945)

In 1945 analyseerde de Britse ecooloog P.H. Leslie een matrixmodel voor een leeftijdsgestructureerde populatie van knaagdieren, waarmee hij het werk van Lotka aanpaste aan een discreet-tijds kader. Hij benadrukte dat de groeisnelheid overeenkomt met een eigenwaarde en de stabiele leeftijdsstructuur met een eigenvector. Hij schatte ook numeriek de netto reproductiesnelheid \mathcal{R}_0 voor de bruine rat.

Patrick Holt Leslie werd in 1900 geboren in de buurt van Edinburgh in Schotland. Hij studeerde aan het *Christ Church College* van de Universiteit van Oxford en behaalde in 1921 een bachelorgraad in de fysiologie. Maar hij kon zijn medische studies niet afmaken wegens gezondheidsproblemen. Na enkele jaren als assistent bacteriologie in de afdeling pathologie te hebben gewerkt, wendde hij zich tot de statistiek en trad in 1935 toe tot het *Bureau of Animal Population*, een nieuw onderzoekscentrum dat door Charles Elton was opgericht. Het doel van dit centrum was de schommelingen van de dierenpopulaties te bestuderen door middel van veldstudies en laboratoriumexperimenten. Het meeste onderzoek werd verricht op knaagdieren: analyse van de cycli van de haas en zijn predator de lynx aan de hand van de archieven van de *Hudson's Bay Company* in Canada, follow-up van de territoriale expansie van de grijze eekhoorn ten koste van de rode eekhoorn in Engeland, gegevensverzameling over woelmuizen in de omgeving van Oxford enzovoort. Leslie paste op de gegevens over woelmuizen de door Lotka ontwikkelde methoden voor de menselijke demografie toe. Tijdens de Tweede Wereldoorlog was het onderzoek van het centrum gericht op methoden voor de bestrijding van ratten en muizen in silo's.

In 1945 publiceerde Leslie zijn beroemdste artikel in *Biometrika*, een tijdschrift dat in 1901 was opgericht door Galton, Pearson en Weldon. Het artikel was getiteld *Over het gebruik van matrices in bepaalde bevolkingswiskunde*. Leslie beschouwde een model voor de groei van het aantal wijfjes in een dierenpopulatie, b.v. een populatie ratten (maar het kunnen ook mensen zijn). De populatie wordt verdeeld in $K + 1$ leeftijdsgroepen: $P_{k,n}$ is het aantal wijfjes van k leeftijd op tijdstip n ($k = 0, 1, \dots, K$; $n = 0, 1, \dots$). Noem f_k de vruchtbaarheid op leeftijd k of nauwkeuriger gezegd het aantal dochters dat per



Figuur 21.1: P. H. Leslie (1900–1972)

vrouwtje geboren wordt tussen tijdstip n en tijdstip $n + 1$. Dan is K de maximumleeftijd met niet-nul vruchtbaarheid ($f_K > 0$). Noem s_k de kans dat een dier met leeftijd k ten minste tot leeftijd $k + 1$ overleeft. De leeftijdsstructuur van de populatie wordt dan gegeven door de volgende reeks vergelijkingen:

$$\begin{cases} P_{0,n+1} = f_0 P_{0,n} + f_1 P_{1,n} + \cdots + f_K P_{K,n} \\ P_{1,n+1} = s_0 P_{0,n} \\ P_{2,n+1} = s_1 P_{1,n} \\ \vdots \\ P_{K,n+1} = s_{K-1} P_{K-1,n} \end{cases}$$

Alle getallen f_k zijn nonnegatief, terwijl s_k voldoet aan $0 < s_k < 1$. Op het einde van de negentiende en in het begin van de twintigste eeuw hadden de wiskundigen de gewoonte om dergelijke stelsels van vergelijkingen te schrijven in de verkorte vorm¹

$$P_{n+1} = M P_n, \quad (21.1)$$

waarin P_n de kolomvector $(P_{0,n}, \dots, P_{K,n})$ is en M de vierkante matrix (d.w.z. de getallentabel met $K + 1$ rijen en $K + 1$ kolommen)

$$M = \begin{pmatrix} f_0 & f_1 & f_2 & \cdots & f_K \\ s_0 & 0 & 0 & \cdots & 0 \\ 0 & s_1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & s_{K-1} & 0 \end{pmatrix}.$$

Om het gedrag van het systeem (21.1) als functie van de tijd te begrijpen, zocht Leslie naar een meetkundig toenemende of afnemende oplossing $P_n =$

¹wat betekent dat $P_{k,n+1} = M_{k,0} P_{0,n} + M_{k,1} P_{1,n} + \cdots + M_{k,K} P_{K,n}$ voor alle k .

$r^n V$. Het getal r en de vector V moeten voldoen aan

$$MV = rV. \quad (21.2)$$

In dit geval wordt r een eigenwaarde en V een eigenvector van de matrix M genoemd. Met andere woorden, het probleem is de leeftijdsverdeling V te vinden die bij elke tijdstap vermenigvuldigd wordt met een constante r . Volgens de terminologie van Lotka worden dergelijke verdelingen stabiel genoemd. Om terug te keren naar de meer gebruikelijke notaties, kan vergelijking (21.2) herschreven worden als

$$\begin{cases} f_0 V_0 + f_1 V_1 + \cdots + f_K V_K = r V_0, \\ s_0 V_0 = r V_1, \quad s_1 V_1 = r V_2, \quad \dots, \quad s_{K-1} V_{K-1} = r V_K. \end{cases}$$

Uit de laatste K vergelijkingen volgt dat

$$V_1 = \frac{s_0 V_0}{r}, \quad V_2 = \frac{s_0 s_1 V_0}{r^2}, \quad \dots \quad V_K = \frac{s_0 s_1 \cdots s_{K-1} V_0}{r^K}.$$

Door dit in de eerste vergelijking te vervangen, te vereenvoudigen met V_0 en te vermenigvuldigen met r^K , verkrijgt Leslie de karakteristieke vergelijking

$$r^{K+1} = f_0 r^K + s_0 f_1 r^{K-1} + s_0 s_1 f_2 r^{K-2} + \cdots + s_0 s_1 \cdots s_{K-1} f_K. \quad (21.3)$$

Dit is een veeltermvergelijking in r van graad $K+1$. Er zijn dus $K+1$ reële of complexe wortels r_1, \dots, r_{K+1} . Bovendien merkt Leslie (met behulp van Descartes' tekenregel voor veeltermen) dat er slechts één reële positieve wortel is. Noem die r_1 .

Leslie suggereerde ook dat, onder de meeste biologisch realistische voorwaarden (die nauwkeurig kunnen worden gemaakt met behulp van de theorie van Perron en Frobenius voor niet-negatieve matrices), de eigenwaarde r_1 strikt groter is dan de modulus van alle andere reële of complexe eigenwaarden (noem ze r_2, \dots, r_{K+1}). Bovendien zijn alle wortels van (21.3) meestal verschillend. Voor elke eigenwaarde r_i kan men een bijbehorende eigenvector vinden. Zij Q de vierkante matrix van grootte $K+1$ waarvan de kolommen $K+1$ de eigenvectoren bevatten die respectievelijk geassocieerd zijn met r_1, \dots, r_{K+1} , dan is $MQ = QD$, waarbij D de diagonaalmatrix $[r_1, \dots, r_{K+1}]$ is. Dus $M = QDQ^{-1}$ en $P_n = M^n P_0 = QD^n Q^{-1} P_0$. Merk op dat D^n de diagonaalmatrix $[(r_1)^n, \dots, (r_{K+1})^n]$ is en dat

$$D^n / r_1^n \longrightarrow \mathcal{D} = [1, 0, \dots, 0]$$

als $n \rightarrow +\infty$ omdat $r_1 > |r_i|$ voor $i \neq 1$. Daarom convergeert $P_n / (r_1)^n$

naar $Q \mathcal{D} Q^{-1} P_0$.

Elke component van de leeftijdsstructuurvector P_n neemt toe of af als $(r_1)^n$. Als $r_1 > 1$, dan neemt de bevolking exponentieel toe. Als $r_1 < 1$, dan neemt ze exponentieel af. Uit vergelijking (21.3) kan men gemakkelijk aantonen dat de voorwaarde $r_1 > 1$ waar is als en slechts als de parameter \mathcal{R}_0 , gedefinieerd door $\mathcal{R}_0 = f_0 + s_0 f_1 + s_0 s_1 f_2 + \cdots + s_0 s_1 \cdots s_{K-1} f_K$, strikt groter is dan 1. Merk op dat $s_0 s_1 \cdots s_{k-1}$ de overlevingskans is tot minstens leeftijd k . De parameter \mathcal{R}_0 is dus het gemiddeld aantal dochters dat uit een vrouwtje geboren wordt gedurende haar leven en is analoog aan de formules (10.2), (12.2) en (16.9). Het huidige model is een soort discrete-tijd analogon van het werk van Lotka (zie hoofdstuk 10) en een veralgemening met leeftijdsafhankelijke vruchtbaarheid van het werk van Euler (zie hoofdstuk 3).

Leslie illustreerde zijn methode aan de hand van door een Amerikaanse collega gepubliceerde gegevens over de vruchtbaarheids- en overlevingscoëfficiënten f_k en s_k voor de bruine rat. Na enkele statistische bewerkingen om de gegevens op een redelijke manier aan te vullen, verkreeg hij $\mathcal{R}_0 \approx 26$.

Leslie's matrixformulering van problemen in de populatiedynamica wordt nu door veel biologen gebruikt. De berekeningen worden sterk vereenvoudigd door moderne computers en wetenschappelijke software die eigenwaarden en eigenvectoren van om het even welke matrix kunnen berekenen. Men kan gemakkelijk zowel de parameter \mathcal{R}_0 als de groeisnelheid r_1 berekenen.

Na de tweede wereldoorlog gebruikte Leslie zijn methode om de groeisnelheid van andere diersoorten te berekenen: vogels, kevers enz. Hij werkte ook aan stochastische modellen, aan modellen van concurrentie tussen soorten en aan de analyse van gegevens die zijn verzameld met de *capture-recapture*-methode. Hij ging in 1967 met pensioen. In datzelfde jaar, nadat Charles Elton ook met pensioen was gegaan, hield het *Bureau of Animal Population* op te bestaan als onafhankelijk onderzoekscentrum en werd het een onderdeel van het Departement Zoölogie van de Universiteit van Oxford. Leslie overleed in 1972.

Verdere lectuur

1. Anonymous: Dr P. H. Leslie. *Nature* 239, 477–478 (1972)
2. Crowcroft, P.: *Elton's Ecologists, A History of the Bureau of Animal Population*. University of Chicago Press (1991)
3. Leslie, P.H.: On the use of matrices in certain population mathematics. *Biometrika* 33, 213–245 (1945)

Hoofdstuk 22

Percolatie en epidemieën (1957)

In 1957 onderzochten Hammersley en Broadbent de verspreiding van een „vloeistof” in een oneindig regelmatig vierkantsnetwerk, waarbij twee naburige knooppunten met een bepaalde kans met elkaar verbonden zijn. Als mogelijke voorbeelden noemden zij de verspreiding van een epidemie in een boomgaard. Zij toonden aan dat er een kritische kans is waaronder geen grote epidemie kan uitbreken en waarboven grote epidemieën met een positieve kans optreden. Hun artikel was de geboorte van de percolatietheorie.

John Michael Hammersley werd in 1920 geboren in Schotland, waar zijn vader werkte voor een Amerikaans bedrijf dat staal exporteerde. Hij ging studeren aan het *Emmanuel College* van de Universiteit van Cambridge, maar moest in 1940 in het leger. Hij werkte aan de verbetering van berekeningen voor de artillerie. Na zijn studie werd hij in 1948 assistent aan de Universiteit van Oxford in de groep die zich bezighield met het ontwerpen en analyseren van experimenten. In 1955 trad hij in dienst bij de *Atomic Energy Research Establishment* in Harwell bij Oxford.



Figuur 22.1:
Hammersley (1920–2004)

Simon Ralph Broadbent werd geboren in 1928. Hij studeerde technische wetenschappen in Cambridge, wiskunde aan het *Magdalen College* in Oxford (waar hij ook poëzie schreef) en begon een doctoraat in de statistiek aan het

Imperial College in Londen met als vertaalde onderwerp *enige tests voor afwijkingen van uniforme spreiding*. Tijdens zijn promotie kreeg hij steun van de Britse Vereniging voor onderzoek naar het gebruik van steenkool, om statistische problemen te onderzoeken die in verband konden worden gebracht met de steenkoolproductie.

In 1954 werd in de *Royal Statistical Society* in Londen een symposium gehouden over Monte Carlo methoden, gesponsord door de *Atomic Energy Research Establishment*. Bij deze methoden, die in de jaren 40 door John von Neumann, Stanisław Ulam en Nicholas Metropolis in het Los Alamos Laboratorium werden ontwikkeld, wordt gebruik gemaakt van stochastische computersimulaties om onbekende wiskundige grootheden te schatten. Hammersley presenteerde op het symposium in Londen een document dat hij had voorbereid in samenwerking met Morton, een collega uit Harwell. Het artikel werd ook gepubliceerd in het *Journal of the Royal Statistical Society*. Tijdens de discussie die volgde op de presentatie op het symposium, noemde Broadbent een interessant probleem dat bestudeerd zou kunnen worden met behulp van een Monte Carlo methode: gegeven een regelmatig netwerk van poriën in twee of drie dimensies zodanig dat twee aangrenzende poriën met een waarschijnlijkheid van p met elkaar verbonden zijn, welk deel van het netwerk zou gevuld worden door een gas als het via een van deze poriën zou worden ingevoerd? Broadbent dacht hierbij aan het ontwerp van gasmaskers voor mijnwerkers en in het bijzonder aan de grootte van de poriën die nodig was voor hun werking.

Hammersley begon toen met Broadbent aan dit gasmaskerprobleem te werken. Zij realiseerden zich dat het slechts een prototype was van een familie van problemen die nog niet waren bestudeerd: de deterministische voortplanting van een „vloeistof” (de betekenis hangt af van de context) in een willekeurig medium. Hammersley noemde het „percolatie”, naar analogie van wat er in een koffiepot gebeurt. Bij de *Atomic Energy Research Establishment* had Hammersley ook toegang tot enkele van de krachtigste computers van zijn tijd om Monte Carlo methoden te testen op percolatie problemen.

In 1957 publiceerden Broadbent en Hammersley eindelijk het eerste artikel over de wiskundige theorie van percolatie. Één van de voorbeelden die zij beschouwden, was een model uit de populatiedynamica, namelijk de verspreiding van een epidemie in een boomgaard. In dit model wordt aangenomen dat de bomen in een zeer grote boomgaard zich op de knopen van een vierkantsrooster bevinden. Elk van de vier dichtstbijzijnde bomen van een bepaalde besmette boom, die niet via een andere weg is besmet, heeft een kans p om door de besmette boom besmet te worden. De vraag is of een groot aantal bomen besmet zal raken of dat de epidemie enkel lokaal zal verspreiden. Dit

hangt natuurlijk af van de kans p , die op zijn beurt weer samenhangt met de afstand tussen de bomen, d.w.z. de breedte van het netwerkraaster.

Broadbent en Hammersley onderzochten het specifieke geval waarin de boomgaard oneindig is en het hele vlak beslaat en er aan het begin van de epidemie slechts één besmette boom is. Laat $f(p)$ de kans zijn dat uiteindelijk een oneindig aantal bomen vanuit deze bron geïnfecteerd raakt. Het is te verwachten dat $f(p)$ een stijgende functie van p is met $f(0) = 0$ en $f(1) = 1$. Hun belangrijkste resultaat was dat er een kritische kans p^* is, $0 < p^* < 1$, zodanig dat:

- als $p < p^*$, dan $f(p) = 0$ dus slechts een eindig aantal bomen is uiteindelijk geïnfecteerd;
- als $p > p^*$, dan $f(p) > 0$ en een oneindig aantal bomen kan geïnfecteerd worden.

Het bewijs wordt geleverd door een vergelijking met het aantal verschillende „zelfvermijdende wandelingen” in het vlak dat begint bij de infectiebron. Deze wandelingen gaan door een bepaald aantal naburige bomen (merk op dat elke boom vier burens heeft) zonder een boom meer dan eens te bezoeken. Een zelfvermijdende wandeling van lengte n is een infectiepad met een waarschijnlijkheid van p^n omdat de infectie van elke bezochte boom naar de volgende kan worden overgedragen met een waarschijnlijkheid van p . Laat nu $q(j, n)$ de kans zijn dat er onder alle zelfvermijdende wandelingen van lengte n met hetzelfde beginpunt, precies j zulke wandelingen zijn die een infectiepad zijn. Als er een oneindig aantal geïnfecteerde bomen is, dan bestaat er voor elk geheel getal n minstens één zelfvermijdende wandeling van lengte n die een infectiepad is. Dus

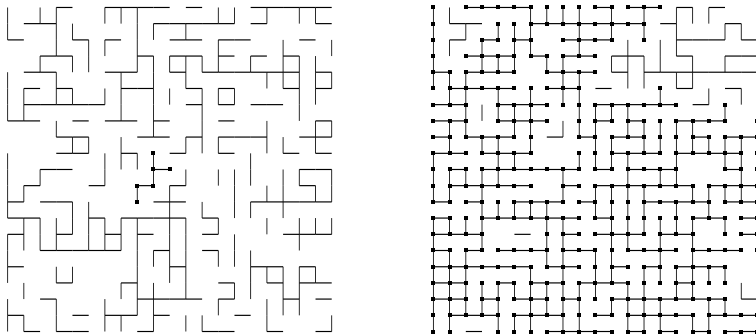
$$0 \leq f(p) \leq \sum_{j=1}^{+\infty} q(j, n) \leq \sum_{j=1}^{+\infty} j q(j, n)$$

voor alle n . Maar $\sum_{j=1}^{+\infty} j q(j, n)$ is het verwachte aantal zelfvermijdende wandelingen van lengte n die infectiepaden zijn. Dit aantal is gelijk aan $p^n s(n)$, waarbij $s(n)$ het totale aantal zelfvermijdende wandelingen van lengte n is. Hammersley kon in een begeleidend artikel aantonen dat $s(n)$ groeit als $e^{\kappa n}$ als $n \rightarrow +\infty$, waarbij κ de verbindingsconstante wordt genoemd. Als $p < e^{-\kappa}$, dan groeit $p^n s(n) \rightarrow 0$ als $n \rightarrow +\infty$ en $f(p) = 0$. Dus $p^* \geq e^{-\kappa} > 0$.

In de praktijk is het dus beter als de bomen niet te dicht bij elkaar staan om p onder p^* te houden in geval van een epidemie. Maar hoe dichter de bomen bij elkaar staan, hoe hoger de productie per hectare. Er moet een compromis worden gevonden.

Zoals Broadbent en Hammersley opmerkten, bestaat er een zekere overeenkomst tussen het bestaan van een kritische kans in percolatieprocessen en het bestaan van een drempelwaarde in vertakkingsprocessen (zie hoofdstuk 7).

Men kan proberen de kritische kans p^* numeriek te schatten. Neem daarvoor een waarde voor p en benader het oneindige netwerk door een eindig vierkantsnetwerk van grootte $N \times N$, waarbij N groot genoeg is. Veronderstel nu dat de boom in het midden van het netwerk geïnfecteerd is. Met een computer kan men willekeurig kiezen welke bomen andere bomen kunnen infecteren. Figuur 22.2 toont de willekeurig gekozen infectiepaden met behulp van verbindingen in een netwerk. In figuur 22.2a is p kleiner dan p^* . In figuur 22.2b is p groter dan p^* . Men kan gemakkelijk bepalen welke bomen geïnfecteerd kunnen worden, namelijk die bomen die bereikt kunnen worden door een pad van verbindingen te beginnen bij de geïnfecteerde boom in het centrum. Deze bomen zijn in de figuren aangegeven met kleine zwarte vierkantjes.

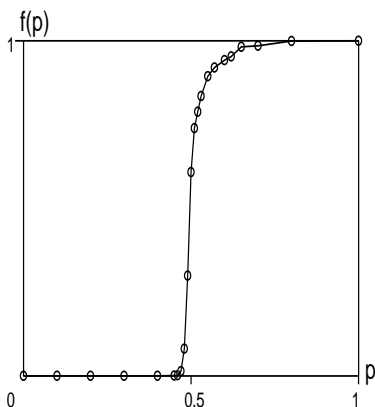


Figuur 22.2: Percolatie met $p = 0,4$ [links] of $p = 0,55$ [rechts].

Men kan dan nagaan of de epidemie de rand van het netwerk van $N \times N$ heeft bereikt. Als dit zo is en als N groot genoeg is, kan men ervan uitgaan dat

het aantal geïnfecteerde bomen „bijna oneindig” is. Door dit soort simulaties vele malen te herhalen, kan men bij benadering de waarde van de kans $f(p)$ vinden dat het aantal geïnfecteerde bomen oneindig is (dit is de Monte Carlo methode). Als men tenslotte p tussen 0 en 1 laat variëren, kan men een benadering van de drempelwaarde p^* krijgen, dat is de kleinste waarde waarvoor geldt dat $f(p) > 0$ indien $p > p^*$.

Het artikel van Broadbent en Hammersley bevatte alleen het bewijs van het bestaan van de drempelwaarde p^* . In de daaropvolgende jaren bleef Hammersley de wiskundige theorie van percolatie verder ontwikkelen, terwijl Broadbent zich op andere onderwerpen toelegde. Met de ontwikkeling van computers in de jaren zeventig werd het eenvoudiger om de hierboven beschreven simulaties uit te voeren (figuur 22.3). Men vermoedde toen dat $p^* = 1/2$. Dit resultaat werd uiteindelijk in 1980 bewezen door Harry Kesten van Cornell-universiteit.



Figuur 22.3: Kans $f(p)$ dat oneindig veel bomen geïnfecteerd raken als functie van p . De curve is verkregen door 1.000 simulaties uit te voeren op een netwerk van 200×200 .

Tussen 1959 en 1969 werkte Hammersley voor het Instituut voor Economie en Statistiek van de Universiteit van Oxford. Hij werd fellow van het *Trinity College*. In 1964 publiceerde hij in samenwerking met David Handcomb een boek getiteld *Monte Carlo Methoden*. In 1976 werd hij verkozen tot lid van de *Royal Society*. Hij ging in 1987 met pensioen maar bleef het Oxford Centrum voor Industriële en Toegepaste Wiskunde bezoeken. Hij overleed in 2004.

Broadbent promoveerde aan *Imperial College* in 1957. Hij vond een baan in een industrieel bedrijf, de *United Glass Bottle Manufacturers*. Na tien

jaar in de industrie ging hij werken bij een persbureau, de *London Press Exchange*, dat wetenschappelijk onderzoek deed naar het lezerspubliek. Het bureau werd in 1969 gekocht door *Leo Burnett*, een Amerikaans reclamebedrijf. Broadbent werkte aan het meten van de effectiviteit van reclame en publiceerde verschillende boeken over dat onderwerp: *Besteding van Reclamegelden* (1975), *Advertentiebudget* (1989), *Verantwoord Reclame* (1997) en *Wanneer te adverteren* (1999). In 1980 was hij medeoprichter van de Reclame Effectiviteit Prijzen (de *Advertising Effectiveness Awards*). Hij werkte enkele jaren op het hoofdkantoor van *Leo Burnett* in Chicago als directeur voor merkeconomie. Hij had ook zijn eigen adviesbureau, *BrandCon Limited*. Hij overleed in 2002.

Verdere lectuur

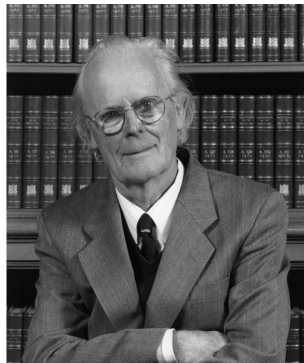
1. Grimmett, G., Welsh, D.: John Michael Hammersley. *Biogr. Mem. Fellows R. Soc.* 53, 163–183 (2007)
2. Broadbent, S.R.: Discussion on symposium on Monte Carlo methods. *J. R. Stat. Soc. B* 16, 68 (1954)
3. Broadbent, S.R., Hammersley, J.M.: Percolation processes I: Crystals and mazes. *Proc. Camb. Philos. Soc.* 53, 629–641 (1957)
4. Broadbent, T.: Simon Broadbent – The man with a sense of fun who gave advertising a value. *Campaign*, 26 April 2002. www.campaignlive.co.uk/news/143366/
5. Hammersley, J.M.: Percolation processes II: The connective constant. *Proc. Camb. Philos. Soc.* 53, 642–645 (1957)
6. Hammersley, J.M.: Percolation processes: lower bounds for the critical probability. *Ann. Math. Stat.* 28, 790–795 (1957)
7. Hammersley, J.M.: Origins of percolation theory. In: Deutscher, G. Zallen, R., Adler, J. (eds.) *Percolation Structures and Processes*, 47–57. Israel Physical Society (1983)
8. Hammersley, J.M., Morton, K.W.: Poor man's Monte Carlo. *J. R. Stat. Soc. B* 16, 23–38 (1954)
9. Hammersley, J.M., Handscomb, D.C.: *Monte Carlo Methods*. Fletcher & Son, Norwich (1964)
10. Kesten, H.: The critical probability of bond percolation on the square lattice equals $1/2$. *Comm. Math. Phys.* 74, 41–59 (1980)
11. Metropolis, N., Ulam, S.: The Monte Carlo method. *J. Amer. Stat. Assoc.* 44, 335–341 (1949)

Hoofdstuk 23

Speltheorie en evolutie (1973)

In 1973 publiceerden Maynard Smith en Price een artikel waarin zij analyseerden waarom dieren vermeden hun gevaarlijkste wapens te gebruiken in intraspecifieke conflicten. Hun model maakte gebruik van de speltheorie en was een van de eerste modellen die deze wiskundige theorie op evolutionaire problemen toepaste.

John Maynard Smith werd in 1920 in Londen geboren. Zijn vader, die chirurg was, overleed toen hij acht jaar oud was. Maynard Smith studeerde aan het *Eton College* en ging ingenieur studeren aan het *Trinity College* van de universiteit van Cambridge. Hij was toen lid van de Communistische Partij van Groot-Brittannië. In 1939, toen de oorlog uitbrak, probeerde hij zich als vrijwilliger bij het leger aan te melden, maar hij werd afgewezen vanwege zijn slechte gezichtsvermogen. Hij voltooide zijn ingenieursstudie en werkte enkele jaren aan het ontwerp van militaire vliegtuigen. Uiteindelijk besloot hij zich toe te leggen op de biologie en studeerde genetica aan het *University College* in Londen met Haldane als promotor. In 1952 werd hij docent in de zoölogie. Hij verliet de Communistische Partij na de gebeurtenissen in Hongarije in 1956. Zijn eerste boek, getiteld *De evolutietheorie*, werd in 1958 gepubliceerd. In 1965 werd hij hoogleraar biologie aan de pas opgerichte Universiteit van Sussex. Daarna publiceerde hij nog twee andere boeken: *Wiskundige Ideeën in de Biologie* (1968) en *Over evolutie* (1972).



Figuur 23.1:
Maynard Smith (1920–2004)

George R. Price werd in 1922 in de Verenigde Staten geboren. Hij studeerde scheikunde aan de Universiteit van Chicago, waar hij in 1946 promoveerde na te hebben meegewerkt aan het Manhattanproject, de bouw van de atoombom. In 1950 werd hij geassocieerd onderzoeker in de geneeskunde aan de Universiteit van Minnesota. Later werkte hij als onafhankelijk journalist voor verschillende tijdschriften voordat hij terugkeerde naar het onderzoek bij IBM. In 1967, na te zijn behandeld voor schildklierkanker, vestigde hij zich in Engeland en wijdde zich aan de studie van een geheel ander onderwerp: evolutiebiologie. Vanaf 1968 werkte hij in Londen in het Galtonlaboratorium van het *University College*. Zijn eerste artikel op dit nieuwe gebied, *Selectie en covariantie*, werd gepubliceerd met de hulp van W. D. Hamilton in een nummer van 1970 van *Nature* en bevatte wat nu de vergelijking van Price wordt genoemd.

Price heeft nog een artikel ingediend bij *Nature*, dit keer over conflicten tussen dieren. Maar het had niet het juiste formaat voor dit tijdschrift. Dus stelde Maynard Smith, die de reviewer was, voor om een kortere versie te maken. Price begon aan iets anders te werken terwijl Maynard Smith het idee van Price zelf begon uit te werken. Uiteindelijk publiceerden Maynard Smith en Price een gezamenlijk artikel, getiteld *De logica van het dierenconflict*, dat *Nature* in 1973 publiceerde. Het artikel leverde een interessante bijdrage aan het gebruik van speltheorie in de evolutiebiologie. Voordien was de speltheorie vooral ontwikkeld voor economie en politiek, vooral na het boek van John von Neumann en Oskar Morgenstern uit 1944, getiteld *Theorie van Spel en Economisch Gedrag*. Het uitgangspunt van Maynard Smith en Price was de volgende vraag: hoe komt het dat in conflicten tussen dieren van dezelfde soort, de „wapens” waarover zij beschikken (hoorns, klauwen, gif enz.) zelden worden gebruikt om te doden? Volgens Darwin’s ideeën over de strijd om het leven, zouden agressievere dieren meer gevechten moeten winnen en een groter aantal nakomelingen moeten krijgen, wat leidt tot een escalatie in het gebruik van „wapens”. Merk op dat dit de tijd van de Koude Oorlog was, dus het onderwerp had ook een politiek tintje.

Maynard Smith en Price stelden zich een opeenvolging van spelen voor waarin twee dieren met elkaar kunnen wedijveren om een hulpbron, bijvoorbeeld een territorium in een gunstige habitat. In de vereenvoudigde voorstelling die Maynard Smith zou gebruiken in zijn boek uit 1982 (*Evolutie en speltheorie*), neemt elk dier ofwel de „havik-strategie” of de „duif-strategie” aan. In wat volgt hebben we het gewoon over haviken en duiven, maar we bedoelen strategieën die door dieren van dezelfde soort worden aangenomen. Stel dat $V > 0$ de waarde van de hulpbron is, wat betekent dat als \mathcal{R}_0 het normale gemiddelde aantal nakomelingen van een dier is, de winnaar van de

competitie gemiddeld $\mathcal{R}_0 + V$ nakomelingen heeft.

Als een havik een andere havik ontmoet, strijden zij om de hulpbron: de winnaar krijgt de hulpbron met waarde V , de verliezer betaalt een „prijs” $C > 0$. Elk van de twee haviken heeft een kans gelijk aan $1/2$ om de strijd te winnen en een even grote kans om te verliezen. De verwachte opbrengst van een gevecht tussen twee haviken is dus $\frac{1}{2}(V - C)$ voor de twee concurrenten. Als een havik echter een duif ontmoet, dan krijgt de havik altijd de hulpbron V terwijl de duif ontsnapt zonder te vechten dus geen prijs betaalt. Tenslotte, als twee duiven elkaar ontmoeten, krijgt één van hen de hulpbron V terwijl de andere ontsnapt zonder te vechten en dus zonder kosten. Aangezien beide duiven evenveel kans hebben om te winnen ($1/2$), is de verwachte opbrengst van een ontmoeting tussen twee duiven dus $V/2$. De uitbetalingen kunnen worden samengevat zoals in Tabel 23.1.

Tabel 23.1: Verwachte uitbetalingen van het havik-duif spel.

	een havik	een duif
de uitbetaling van een havik tegen...	$\frac{1}{2}(V - C)$	V
de uitbetaling van een duif tegen...	0	$V/2$

Meer algemeen kan men zich gevechten voorstellen tussen individuen die één van twee strategieën kunnen aannemen, noem ze 1 en 2, met een matrix van verwachte uitbetalingen $(G_{i,j})_{1 \leq i,j \leq 2}$. In bovenstaand voorbeeld volgen haviken strategie 1, duiven strategie 2, $G_{1,1} = \frac{1}{2}(V - C)$, $G_{1,2} = V$, $G_{2,1} = 0$ en $G_{2,2} = V/2$. In het oorspronkelijke artikel van 1973 hadden Maynard Smith en Price reeds computersimulaties gebruikt om meer dan twee mogelijke strategieën te testen (deze werden „havik”, „muis”, „pester”, „wreker” en „tester-wreker” genoemd).

Stelt u zich nu een grote populatie dieren van dezelfde soort voor met in generatie n een aandeel haviken van x_n en een aandeel duiven van $1 - x_n$. De haviken in generatie n hebben een gemiddeld aantal nakomelingen dat gelijk is aan

$$R_1(n) = \mathcal{R}_0 + x_n G_{1,1} + (1 - x_n) G_{1,2}. \quad (23.1)$$

Evenzo hebben duiven een gemiddeld aantal nakomelingen gelijk aan

$$R_2(n) = \mathcal{R}_0 + x_n G_{2,1} + (1 - x_n) G_{2,2}. \quad (23.2)$$

Het gemiddelde aantal nakomelingen in de hele populatie is dus

$$R(n) = x_n R_1(n) + (1 - x_n) R_2(n).$$

Zonder rekening te houden met mogelijke subtiliteiten als gevolg van seksuele voortplanting, zien we dat het aandeel haviken in de volgende generatie

$$x_{n+1} = x_n R_1(n) / R(n). \quad (23.3)$$

Bijgevolg is $x_{n+1} > x_n$ als $R_1(n) > R(n)$ en $x_{n+1} < x_n$ als $R_1(n) < R(n)$. Er zijn drie mogelijke stationaire toestanden: $x = 0$, $x = 1$ en

$$x^* = \frac{G_{1,2} - G_{2,2}}{G_{2,1} - G_{1,1} + G_{1,2} - G_{2,2}}$$

op voorwaarde dat $0 < x^* < 1$. In het havik-duif spel,

$$x^* = V/C < 1$$

op voorwaarde dat $V < C$.

Inderdaad, $x = 0$ is een voor de hand liggende stationaire toestand van (23.3). Als $x = 0$ een andere stationaire toestand is, dan is $R_1 = R = xR_1 + (1-x)R_2$. Dus ofwel $x = 1$ ofwel $R_1 = R_2$. De laatste mogelijkheid is equivalent met $xG_{1,1} + (1-x)G_{1,2} = xG_{2,1} + (1-x)G_{2,2}$, wat de stationaire toestand x^* oplevert.

De stationaire toestand $x = 1$ komt overeen met een populatie waarvan 100% van de individuen strategie 1 volgt. Deze stationaire toestand is stabiel als hij niet kan worden geïnvadeerd door enkele individuen die strategie 2 volgen. Uit (23.3) blijkt dat deze voorwaarde gelijk is aan $R_1(n) > R(n)$ voor alle x_n die voldoende dicht bij 1 liggen. Aangezien $R(n) = x_n R_1(n) + (1-x_n)R_2(n)$ wordt de voorwaarde $R_1(n) > R_2(n)$ voor alle x_n die voldoende dicht bij 1 liggen. Als we de uitdrukkingen (23.1)-(23.2) van R_1 en R_2 bekijken, komen we tot de conclusie dat $x = 1$ stabiel is als en slechts als aan een van de twee volgende voorwaarden is voldaan:

- $G_{1,1} > G_{2,1}$;
- $G_{1,1} = G_{2,1}$ en $G_{1,2} > G_{2,2}$.

Indien dit het geval is, wordt strategie 1 een „evolutionair stabiele strategie” genoemd¹. In het havik-duif spel is de voorwaarde $G_{1,2} > G_{2,2}$ altijd waar. De havik-strategie is dus evolutionair stabiel alleen wanneer $G_{1,1} \geq G_{2,1}$, d.w.z. wanneer $V \geq C$.

¹Preciezer gezegd, de stationaire toestand $x = 1$ is een „evolutionair stabiele toestand”. Het onderscheid tussen toestand en strategie is belangrijk wanneer er meer dan twee strategieën zijn.

De stationaire toestand $x = 0$ komt overeen met een populatie waarvan alle individuen strategie 2 volgen. Deze situatie is symmetrisch met de vorige als we de indices 1 en 2 verwisselen. In het havik-duif spel hebben we $G_{1,2} = V > G_{2,2} = V/2$ dus de stationaire toestand $x = 0$ is altijd onstabiel. De introductie van een klein aantal haviken in een populatie duiven zou leiden tot een geleidelijke invasie van de haviken.

Op dezelfde manier kan men aantonen dat de derde stationaire toestand x^* , op voorwaarde dat $0 < x^* < 1$, altijd stabiel is. In het havik-duifspel komt $x^* = V/C$ overeen met een gemengde populatie met zowel haviken als duiven.

Samenvattend zijn er twee gevallen in het havik-duif spel. Als de waarde van de hulpbron groter is dan de mogelijke kosten, dan neigt de populatie naar een evenwicht met alleen haviken en geen duiven, ongeacht de begintoestand $x(0)$ met $0 < x(0) < 1$. De havikstrategie is dan een evolutionair stabiele strategie. Indien daarentegen $V < C$, dan neigt de populatie naar een gemengde steady state met een aandeel x^* haviken en een aandeel $1 - x^*$ duiven. Het model geeft dus een verklaring waarom individuen met minder agressief gedrag kunnen overleven wanneer $V < C$. De formule $x^* = V/C$ laat bovendien zien dat hoe hoger de kosten C voor verliezers zijn, hoe kleiner het aandeel x^* van haviken in de populatie is. Vandaar dat soorten met de gevaarlijkste „wapens” deze zelden gebruiken voor intraspecifieke gevechten: zij geven de voorkeur aan onoffensieve rituele gevechten, waarbij concurrerende dieren elkaar proberen te imponeren, maar echte gevechten die verwondingen kunnen veroorzaken, vermijden.

In het oorspronkelijke artikel van Maynard Smith en Price uit 1973 werd het concept van evolutionair stabiele strategie besproken en werd voornamelijk gebruik gemaakt van computersimulaties van conflicten tussen dieren, waarbij de uitbetalingen van verschillende strategieën werden geregistreerd. De benadering met dynamische vergelijkingen zoals (23.3) werd iets later ontwikkeld, met name door Taylor en Jonker. Sindsdien hebben veel auteurs ideeën uit de speltheorie toegepast op vragen in de evolutiebiologie of omgekeerd dynamische evolutionaire benaderingen toegepast op meer klassieke problemen in de speltheorie. Naast vragen betreffende conflicten tussen dieren, kan men bijvoorbeeld problemen van ouderlijke investering of van geslachtsverhouding (de verhouding tussen het aantal mannetjes en vrouwtjes bij de geboorte) noemen, welke laatste reeds door Carl Düsing in 1884 en door Ronald Fisher in zijn boek van 1930 over *De Genetische Theorie van Natuurlijke Selectie* werden bestudeerd. Sommige andere modellen richten zich op de dynamische aspecten van het „gevangenendilemma” of van het „steen-papier-schaar”-spel. Men realiseerde zich ook dat het concept van evolutionair stabiele strategie nauw verwant is aan het concept van Nash-

evenwicht in de speltheorie.

Price, die een overtuigd atheïst was geweest, had in 1970 een mystieke ervaring en bekeerde zich tot het christelijk geloof. In 1974 gaf hij zijn onderzoek op omdat hij vond dat „het soort theoretische wiskundige genetica waar hij mee bezig was, niet erg relevant was voor menselijke problemen”. Hij gaf al zijn bezittingen aan daklozen en pleegde een paar maanden later zelfmoord.

Maynard Smith daarentegen zette deze gedachtengang voort en werd in 1977 tot lid van de *Royal Society* gekozen. Hij publiceerde vele boeken: *Modellen in Ecologie* (1974), *De evolutie van seks* (1978), *Evolutie en de theorie van het spel* (1982), *De problemen van de biologie* (1986), *Had Darwin het bij het rechte eind?* (1988) en *Evolutionaire Genetica* (1989). Hij publiceerde ook in samenwerking met E. Szathmáry *De grote overgangen in de evolutie* (1995) en *De oorsprong van het leven: Van het ontstaan van het leven tot de oorsprong van de taal* (1999). Hij ging in 1985 met pensioen. In 1999 ontving hij de Crafoord prijs in biowetenschappen van de Koninklijke Zweedse Academie van Wetenschappen voor zijn „fundamentele bijdragen aan de conceptuele ontwikkeling van de evolutionaire biologie”. In 2003 publiceerde hij in samenwerking met D. Harper *Dierlijke signalen*. Hij overleed in Sussex in 2004.

Verdere lectuur

1. Charlesworth, B., Harvey, P.: John Maynard Smith, 6 January 1920–19 April 2004. *Biog. Mem. Fellows R. Soc.* 51, 253–265 (2005)
2. Edwards, A.W.F.: Carl Düsing (1884) on the regulation of the sex-ratio. *Theor. Pop. Biol.* 58, 255–257 (2000)
3. Frank, S.A.: George Price’s contributions to evolutionary genetics. *J. Theor. Biol.* 175, 373–388 (1995)
4. Maynard Smith, J., Price, G.R.: The logic of animal conflict. *Nature* 246, 15–18 (1973)
5. Maynard Smith, J.: *Evolution and the Theory of Games*. Cambridge University Press (1982)
6. Schwartz, J.: Death of an altruist: Was the man who found the selfless gene too good for this world? *Lingua Franca* 10, 51–61 (2000) bio.kuleuven.be/ento/pdfs/schwartz2000.pdf
7. Sigmund, K.: John Maynard Smith and evolutionary game theory. *Theor. Pop. Biol.* 68, 7–10 (2005)
8. Taylor, P.D., Jonker, L.B.: Evolutionary stable strategies and game dynamics. *Math. Biosci.* 40, 145–156 (1978)
9. Von Neumann, J., Morgenstern, O.: *Theory of Games and Economic Behavior*. Princeton University Press (1944) archive.org

Hoofdstuk 24

Chaotische populaties (1974)

In 1974 bestudeerde Robert May, een Australische natuurkundige die ecooloog werd, de logistische vergelijking in discrete tijd als een model voor de dynamica van populaties. Hij merkte dat er onverwachte bifurcaties optraden en dat het asymptotische gedrag zelfs chaotisch kon zijn. Voorspellingen op lange termijn kunnen dus zelfs op basis van een eenvoudig deterministisch model onmogelijk zijn. Het artikel van May was één van de artikelen die de „chaostheorie” lanceerden.

Robert McCredie May werd in 1936 in Australië geboren. Na theoretische natuurkunde te hebben gestudeerd en in 1959 te zijn gepromoveerd aan de Universiteit van Sydney, werkte hij twee jaar op de afdeling toegepaste wiskunde van de Harvard-universiteit. Terug in Australië werd hij professor in de theoretische fysica. In 1971, tijdens een bezoek aan het *Institute for Advanced Study* in Princeton, veranderde hij zijn onderzoeksonderwerp en ging hij zich toeleggen op de dynamica van dierenpopulaties. In 1973 werd hij hoogleraar zoölogie in Princeton. In datzelfde jaar publiceerde hij een boek getiteld *Stabiliteit en complexiteit in model-ecosystemen*.



Figuur 24.1:
Robert M. May (1936-2020)

In 1974 publiceerde May in *Science* een artikel getiteld *Biologische populaties met niet-overlappende generaties: stabiele punten, stabiele cycli en chaos*, waarin hij aantoonde dat zeer eenvoudige wiskundige modellen in populatiedynamica zich op chaotische wijze kunnen gedragen.

Om de oorsprong van dit probleem te begrijpen, moet men ongeveer tien jaar teruggaan in de tijd. In 1963 had Edward Lorenz, een Amerikaans me-

teoroloog die aan het *Massachusetts Institute of Technology* (M.I.T.) werkte, bij het maken van numerieke simulaties op zijn computer opgemerkt dat een vereenvoudigd model van de atmosfeer met slechts drie differentiaalvergelijkingen zich op een zeer verrassende manier kon gedragen: een minieme verandering van de beginvoorwaarden kon het eindresultaat van een simulatie en dus ook de weersvoorspellingen volledig veranderen. De wiskundige Henri Poincaré, die zich had beziggehouden met de beweging van planeten in het zonnestelsel, had in feite reeds aan het begin van de twintigste eeuw, lang vóór het computertijdperk, aan deze mogelijkheid gedacht. Maar in het begin van de jaren zeventig begonnen enkele onderzoekers deze vreemde eigenschap nader te bestuderen. Aan de universiteit van Maryland dacht James Yorke na over het werk van Lorenz en introduceerde in dit verband de term „chaos”. Het artikel¹ dat hij samen met zijn student Tien-Yien Li schreef, getiteld *Periode drie impliceert chaos*, verscheen in 1975.

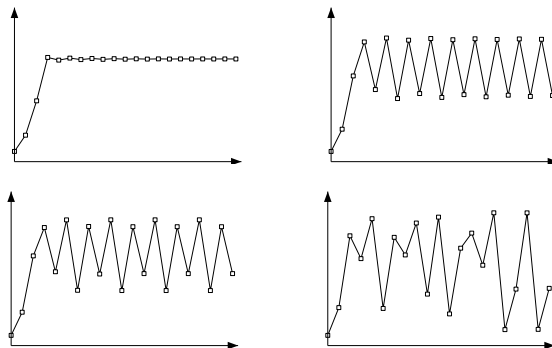
Op zijn beurt bestudeerde May het model

$$p_{n+1} = p_n + a p_n(1 - p_n/K), \quad (24.1)$$

waarbij a en K positieve parameters zijn en p_n staat voor de grootte van een dierenpopulatie in jaar n . Wanneer p_n klein is in vergelijking met de draagkracht K , benadert de dynamiek een geometrische groei $p_{n+1} \approx (1 + a) p_n$. De volledige vergelijking is een soort discrete-tijd analogon van de logistische vergelijking geïntroduceerd door Verhulst (zie hoofdstuk 6). Maar in tegenstelling tot deze laatste vergelijking toonde May aan dat de discrete-tijd versie een veel verrassender gedrag kan vertonen. Bovendien kan dat gedrag makkelijk worden waargenomen met een eenvoudige zakrekenmachine die kan optellen en vermenigvuldigen (figuur 24.2). Maynard Smith had de vergelijking (24.1) reeds bekeken in zijn boek uit 1968 *Wiskundige ideeën in de biologie*. Maar ondanks het feit dat hij verschillende waarden voor a had uitgetoetst, had hij niet ingezien dat er iets bijzonders gaande was.

In figuur p_n , die identiek is aan een figuur in het artikel van May uit 1974, is te zien dat de populatie p_n convergeert naar een stationaire toestand als $0 < a < 2$. Wanneer $2 < a \leq 2,449$ (de bovengrens 2,449 is een benadering), gaat de populatie p_n naar een cyclus met periode 2. Als $2,450 \leq a \leq 2,544$, neigt de populatie p_n naar een cyclus met periode 4. Wanneer $2,545 \leq a \leq 2,564$ heeft p_n een cyclus van periode 8. Enzovoort. De intervallen van de parameter a waarvoor p_n naar een cyclus met periode 2^n neigt, worden kleiner naarmate n toeneemt en overschrijden nooit 2,570. Wanneer $a \geq 2,570$, kan p_n zich op een „chaotische” manier gedragen.

¹Opmerkelijk genoeg werd een meer algemeen resultaat bewezen door O. M. Sharkovskij in 1964, maar zijn artikel gepubliceerd in een Oekraïens wiskundetijdschrift was niet erg bekend.



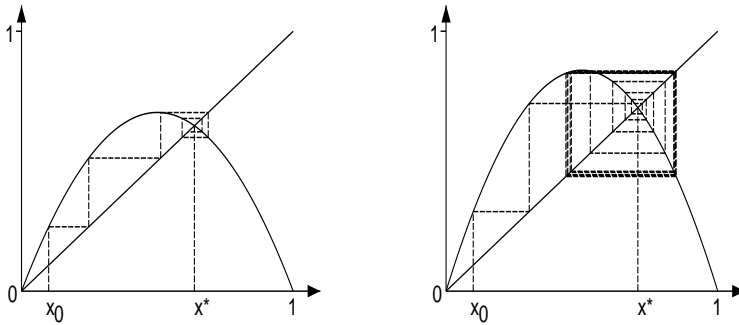
Figuur 24.2: In alle figuren: n staat op de horizontale as, p_n op de verticale as en $p_0 = K/10$. De lijnen worden verkregen door de punten met coördinaten (n, p_n) te verbinden. Linksboven: $0 < a < 2$ (stationaire toestand). Rechtsboven: $2 < a \leq 2,449$ (cyclus van periode 2). Linksonder: $2,450 < a \leq 2,544$ (cyclus van periode 4). Rechtsonder: $2,570 \leq a \leq 3$ (mogelijk chaos).

In 1976 schreef May een overzicht van het probleem, gepubliceerd in *Nature* en getiteld *Eenvoudige wiskundige modellen met een zeer ingewikkelde dynamiek*. Daarin verzamelde hij niet alleen zijn eigen resultaten, maar ook die van andere onderzoekers. Ten eerste, als we $x_n = \frac{ap_n}{K(1+a)}$ en $r = 1 + a$ (zodat $r > 1$) stellen, zien we dat vergelijking (24.1) kan worden herschreven in de eenvoudigere vorm

$$x_{n+1} = rx_n(1 - x_n). \quad (24.2)$$

Opdat deze vergelijking betekenis zou hebben in de populatiedynamica, moet x_n niet-negatief zijn voor alle n . We nemen dus aan dat de beginvoorwaarde x_0 voldoet aan $0 \leq x_0 \leq 1$ en dat $r \leq 4$ is. Deze laatste voorwaarde zorgt ervoor dat het rechterlid van (24.2) tussen 0 en 1 blijft. Opmerkelijk is dat het chaotische geval $r = 4$ reeds in 1947 door Stanisław Ulam en John von Neumann als toevalsgenerator werd gebruikt. Als we de volgende functie invoeren $f(x) = rx(1 - x)$, dan kan vergelijking (24.2) herschreven worden als $x_{n+1} = f(x_n)$ en de stationaire toestanden zijn de oplossingen van $x = f(x)$. Grafisch zijn dit de snijpunten van de krommen $y = f(x)$ en $y = x$ (figuur 24.3). Merk op dat $x = 0$ altijd een stationaire toestand is. Aangezien $r > 1$, bestaat er ook een andere stationaire toestand $x^* > 0$ zodat $x^* = rx^*(1 - x^*)$, d.w.z. $x^* = 1 - 1/r$.

Omdat $r > 1$, is de stationaire toestand $x = 0$ onstabiel. Als x_n dicht bij



Figuur 24.3: De functie $y = f(x)$, de rechte $y = x$, de stationaire toestand x^* en de reeks gedefinieerd door $x_{n+1} = f(x_n)$. (a) $r = 2,75$, de reeks neigt naar x^* . (b) $r = 3,4$, de stationaire toestand x^* is onstabiel en de reeks neigt naar een cyclus met periode 2.

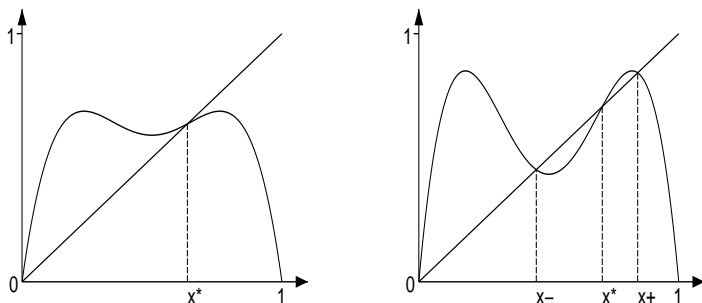
0 ligt, hebben we immers $x_{n+1} \approx r x_n$. Dus x_n neigt weg van 0. Wat de stationaire toestand x^* betreft, deze is enkel lokaal stabiel voor $1 < r < 3$.

Inderdaad, neem $y_n = x_n - x^*$. Dan is (24.2) equivalent met $y_{n+1} = (2 - r - r y_n) y_n$. Als x_n dicht bij x^* ligt, dan ligt y_n dicht bij 0 en $y_{n+1} \approx (2 - r) y_n$. Maar als $y_{n+1} = k y_n$, dan $y_n = k^n y_0$ zodat $y_n \rightarrow 0$ voor $n \rightarrow \infty$ als en slechts als $-1 < k < 1$. Dus de stationaire toestand x^* is lokaal stabiel als en slechts als $-1 < 2 - r < 1$, dus indien $1 < r < 3$.

Wanneer $1 < r < 3$, kan men aantonen dat voor alle beginvoorwaarden $0 < x_0 < 1$, de reeks x_n naar x^* neigt (figuur 24.3a). Maar wat gebeurt er als $3 < r \leq 4$? Om deze vraag te beantwoorden, stellen we eerst vast dat $x_{n+2} = f(x_{n+1}) = f(f(x_n))$. We voeren de volgende functie in, $f_2(x) = f(f(x)) = r^2 x(1-x)[1-rx(1-x)]$ en we beschouwen de oplossingen van de vergelijking $x = f_2(x)$, die dekpunten van de functie $f_2(x)$ genoemd worden. Grafisch zijn dit de snijpunten van de krommen $y = f_2(x)$ en $y = x$ (figuur 24.4).

Als $x = f(x)$, dan is $x = f(f(x)) = f_2(x)$. Dus $x = 0$ en $x = x^*$ zijn ook dekpunten van de functie $f_2(x)$. Maar als $r > 3$ heeft de functie $f_2(x)$ nog twee andere dekpunten, x_- en x_+ , zodat $f(x_-) = x_+$ en $f(x_+) = x_-$.

We zien inderdaad dat $f_2'(x) = f'(f(x)) f'(x)$ zodat $f_2'(x^*) = [f'(x^*)]^2$. Maar $f'(x) = r(1-2x)$ en $x^* = 1-1/r$. Dus $f'(x^*) = 2-r$ en $f_2'(x^*) =$



Figuur 24.4: De krommen $y = f_2(x) = f(f(x))$ en $y = x$ en de stationaire toestand x^* . (a): $r = 2,75$. (b): $r = 3,4$ en de twee andere oplossingen x_- en x_+ van vergelijking $x = f_2(x)$.

$(2-r)^2$. Vandaar dat de helling van de functie $f_2(x)$ op $x = x^*$ zodanig is dat $f_2'(x^*) > 1$ als $r > 3$. Maar aangezien $f_2(0) = 0$, $f_2'(0) = r^2 > 1$ en $f_2(1) = 0$, zien we in figuur 24.4b dat er noodzakelijkerwijs twee andere oplossingen x_- en x_+ zijn van vergelijking $x = f_2(x)$, met $0 < x_- < x^*$ en $x^* < x_+ < 1$. Een andere manier om tot dezelfde conclusie te komen bestaat uit het oplossen van vergelijking $x = f_2(x)$, een veeltermvergelijking van graad 4 met twee bekende wortels: $x = 0$ en $x = x^*$. De twee andere oplossingen x_- en x_+ zijn de wortels van de veelterm

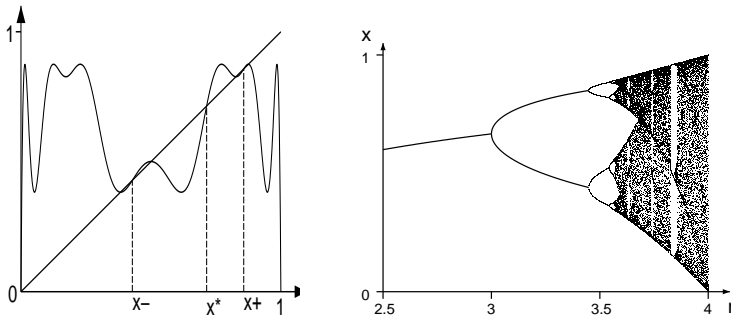
$$x^2 - \frac{1+r}{r}x + \frac{1+r}{r^2} = 0. \quad (24.3)$$

De wortels zijn reëel als de discriminant positief is, d.w.z. als $r > 3$. Omdat $f_2(f(x_-)) = f(f(f(x_-))) = f(f_2(x_-)) = f(x_-)$, is het punt $f(x_-)$ ook een vast punt van $f_2(x)$. Maar $f(x_-) \neq x_-$ want x_- is geen vast punt van $f(x)$. En $f(x_-) \neq x^*$, anders zou $x_- = f(f(x_-)) = f(x^*) = x^*$ zijn. Aangezien $f(x_-) \neq 0$, concluderen we dat $f(x_-) = x_+$. Evenzo geldt dat $f(x_+) = x_-$.

Voor $r > 3$ zien we dus dat als bijvoorbeeld $x_0 = x_-$, dan $x_1 = x_+$, $x_2 = x_-$, $x_3 = x_+$, enz. Men kan ook aantonen dat voor bijna elke beginvoorwaarde $0 < x_0 < 1$ de rij x_n voor $n \rightarrow +\infty$ naar de cyclus van periode 2 neigt, x_-, x_+, x_-, x_+ , enz. (figuur 24.3b en 24.4b). Deze cyclus blijft stabiel zolang r onder de kritische waarde $r_1 = 1 + \sqrt{6} \approx 3,449$ ligt, waar $f_2'(x_-) = -1$.

Inderdaad zien we met behulp van (24.3) dat $f_2'(x_-) = f'(f(x_-))f'(x_-) = f'(x_+)f'(x_-) = r^2(1-2x_+)(1-2x_-) = r^2(1-2(x_++x_-)+4x_+x_-) = r^2\left(1-2\frac{1+r}{r}+4\frac{1+r}{r^2}\right) = -r^2+2r+4$. Dus $f_2'(x_-) = -1$ als $-r^2+2r+5=0$ en in het bijzonder als $r = 1 + \sqrt{6}$.

Voor $r_1 < r < r_2$ wordt een cyclus van periode 4 stabiel: er verschijnen vier nieuwe vaste punten van de functie $f_4(x) = f_2(f_2(x)) = f(f(f(f(x))))$ (figuur 24.5a). Voor $r_2 < r < r_3$ is het een cyclus van lengte 8, enz. De getallen r_n neigen naar een limiet $r_\infty \approx 3,570$ als $n \rightarrow +\infty$. Wanneer $r_\infty < r \leq 4$, kan het systeem zelfs chaotisch zijn! Figuur 24.5b toont het bifurcatiediagram², dat een idee geeft van de complexiteit van de dynamiek.



Figuur 24.5: (a) De kromme $y = f_4(x)$ als $r = 3,5$ en de rechte $y = x$. Naast x^* , x_+ en x_- zijn er nog vier andere vaste punten, die niet gemakkelijk te onderscheiden zijn. (b) Bifurcatiediagram van vergelijking (24.2).

R. M. May besloot met te benadrukken dat zelfs zeer eenvoudige dynamische systemen een zeer gecompliceerd gedrag kunnen vertonen. Dit is niet specifiek voor de vergelijking $x_{n+1} = rx_n(1-x_n)$. Dezelfde „periodeverdubbelingscascade” die tot chaos leidt komt ook voor bij andere vergelijkingen met een functie $f(x)$ die de vorm heeft van een „bult”. Dit is bijvoorbeeld het geval met een andere vergelijking die in de populatiebiologie wordt gebruikt: $x_{n+1} = x_n \exp(r(1-x_n))$.

²Dit diagram werd verkregen door voor elke gegeven waarde van r de punten met coördinaten $(r, x_{200}), (r, x_{201}), (r, x_{220})$ te plotten, waarbij $x_{n+1} = f(x_n)$ en $x_0 = 0, 1$. Als x_n naar een stationaire toestand neigt, zien we slechts één punt in het diagram. Als x_n naar een cyclus van periode 2 neigt, zien we twee punten, enz.

Deze studie suggereert dat men niet verbaasd moet zijn als vele tijdreeksen in populatiedynamica moeilijk te analyseren zijn. Het model toont ook aan dat het onderscheid tussen deterministische en stochastische modellen niet zo duidelijk is als vroeger werd gedacht: zelfs met een eenvoudig deterministisch model kan het onmogelijk zijn lange-termijnvoorspellingen te doen als de parameters zich in het chaotische regime bevinden.

In 1979 werd May verkozen tot lid van de *Royal Society*. Van 1988 tot 1995 was hij hoogleraar aan de Universiteit van Oxford en aan het *Imperial College* in Londen. Van 1995 tot 2000 was hij wetenschappelijk hoofadviseur van de Britse regering. In 1996 ontving hij de Crafoordprijs „voor zijn baanbrekend ecologisch onderzoek met betrekking tot de theoretische analyse van de dynamiek van populaties, gemeenschappen en ecosystemen”. Van ecologie richtte hij zich op epidemiologie en immunologie en publiceerde twee boeken: *Infectieziekten van de mens* (1991, met Roy Anderson) en *Virusdynamica, De wiskundige grondslagen van immunologie en virologie* (2000, met Martin Nowak). In dit laatste boek wordt de interactie tussen de cellen van het immuunsysteem en HIV (het virus dat AIDS veroorzaakt) geanalyseerd als een soort roofdier-prooi systeem (zie hoofdstuk 13). Van 2000 tot 2005 was May voorzitter van de *Royal Society*. In 1996 werd hij geridderd en in 2001 werd hij benoemd tot *life peer*. Hij stierf in 2020.

Verdere lectuur

1. Gleick, J.: *Chaos, Making a New Science*. Viking Penguin, New York (1987)
2. Levin, S.A.: Robert May receives Crafoord prize. *Not. Amer. Math. Soc.* 43, 977–978 (1996) ams.org
3. Li, T.Y., Yorke, J.A.: Period three implies chaos. *Amer. Math. Monthly* 82, 985–992 (1975)
4. Lorenz, E.N.: Deterministic nonperiodic flow. *J. Atmosph. Sci.* 20, 130–141 (1963) journals.ametsoc.org
5. May, R.M.: Biological populations with nonoverlapping generations: stable points, stable cycles and chaos. *Science* 186, 645–647 (1974)
6. May, R.M.: Simple mathematical models with very complicated dynamics. *Nature* 261, 459–467 (1976)
7. May, R.M., Oster, G.F.: Bifurcations and dynamic complexity in simple ecological models. *Amer. Natur.* 110, 573–599 (1976)
8. Maynard Smith, J.: *Mathematical Ideas in Biology*. Cambridge (1968)
9. Poincaré, H.: *Science et Méthode*. Flammarion, Paris (1908) gallica.bnf.fr
10. Sharkovsky, O.M.: Co-existence of cycles of a continuous mapping of a line onto itself. *Ukr. Math. J.* 16, 61–71 (1964)
11. Ulam, S.M., von Neumann, J.: On combination of stochastic and deterministic processes. *Bull. Amer. Math. Soc.* 53, 1120 (1947) ams.org

Hoofdstuk 25

China's één-kind politiek (1980)

In 1980 berekenden Song Jian en zijn medewerkers, die gespecialiseerd waren in controletheorie toegepast op ruimtevaart, dat indien het geboortecijfer in China op het huidige peil zou blijven, de bevolking in de eenentwintigste eeuw meer dan twee miljard zou bereiken. Hun resultaten, die gebaseerd waren op een wiskundig model met leeftijdsstructuur, droegen bij tot het besluit van de regering om over te gaan op een één-kind-beleid.

Song Jian¹ werd in 1931 geboren in Rongcheng in de Chinese provincie Shandong. In de jaren vijftig studeerde hij in de Sovjet-Unie aan de Technische Staatsuniversiteit van Moskou Bauman en aan het Departement Wiskunde en Mechanica van de Staatsuniversiteit van Moskou. Daarna keerde hij terug naar China en werd hoofd van het Bureau voor Cybernetisch Onderzoek in het Instituut voor Wiskunde van de Chinese Academie van Wetenschappen. Hij was een specialist in de toepassing van de controletheorie op de geleiding van raketten. Hij werkte ook voor het Ministerie van de Zevende Machinebouw, dat later werd omgedoopt tot Ministerie van Lucht- en Ruimtevaart. In 1978 begon hij zich te richten op de verbanden tussen controletheorie en demografie.



Figuur 25.1: Song Jian

Om de context van Song Jian's werk over populatiedynamica te begrijpen, moet men eerst een idee geven van wat „controletheorie” is. Het is de

¹Song is de familienaam. Die wordt altijd eerst in het Chinees geschreven.

studie van dynamische systemen waarvan het gedrag afhangt van enkele parameters die in de loop van de tijd kunnen worden gewijzigd om een bepaald criterium te optimaliseren. Deze theorie is met name ontwikkeld in verband met ruimtevaartprogramma's in de VS en in de USSR. Ingenieurs moesten immers de baan van space shuttles „controleren” om satellieten in hun baan rond de aarde te brengen. Maar de toepassingen beperkten zich niet tot natuurkundige of technische problemen. Het geboortebeperkingsbeleid kan ook worden beschouwd als een soort optimaal controleprobleem in wiskundige zin.

Vermeldenswaard is ook het essay getiteld *De grenzen aan de groei: Een rapport voor de Club van Rome's Projekt over de Onzekerheid van de Mensheid*, gepubliceerd in 1972 en geschreven door een groep van het *Massachusetts Institute of Technology* (M.I.T.). Deze studie was gebaseerd op een wiskundig model van de economische groei in de wereld, waarbij rekening werd gehouden met natuurlijke hulpbronnen, bevolkingsomvang en vervuiling. Het rapport suggereerde dat de wereldeconomie afstevende op een catastrofe door uitputting van niet-hernieuwbare hulpbronnen, door gebrek aan voedsel voor de bevolking of door een overmaat aan vervuiling. De vrijwillige beperking van het aantal geboorten was een van de voorgestelde oplossingen. Samengevat was het een soort moderne versie van de stellingen van Malthus. Het rapport kreeg een grote weerklank in het Westen in de jaren 1970.

Sinds de oprichting van de Volksrepubliek in 1949 was het Chinese geboortecijfer zeer hoog, behalve tijdens de catastrofale „Grote sprong voorwaarts”. Halverwege de jaren zeventig herstelde China zich langzaam van de Culturele Revolutie. Gezinsplanning spoorde vrouwen aan geboorten uit te stellen, de tijd tussen twee opeenvolgende geboorten te verlengen en minder kinderen te krijgen. Deng Xiaoping, die na de dood van Mao Zedong in 1976 de nieuwe leider werd, begon in 1978 met het beleid van „vier modernisering”: landbouw, industrie, wetenschap en technologie, en nationale defensie. De omvang en de groei van de Chinese bevolking werden toen gezien als belangrijke obstakels voor deze modernisering. Wetenschappers die tot dan toe aan militaire toepassingen hadden gewerkt, werden aangemoedigd om oplossingen voor dit moeilijke probleem te vinden.

Met deze achtergrond ging Song Jian in 1978 naar Helsinki voor een congres van de *International Federation of Automatic Control*. Hij merkte daar dat sommige onderzoekers in Europa hadden geprobeerd de controletheorie toe te passen op bevolkingsproblemen met het idee dat een strikte geboortebeperving uiteindelijk de catastrofes zou kunnen voorkomen die in het rapport over *De grenzen aan de groei* waren aangekondigd. Terug in China richtte hij

een klein team op, bestaande uit zijn collega Yu Jingyuan en de computerdeskundige Li Guangyuan, om dit soort wiskundige modellering toe te passen op gegevens betreffende de Chinese bevolking. In die tijd was er nauwelijks wetenschappelijke communicatie tussen China en de rest van de wereld. Het team herontwikkelde de vergelijkingen die de evolutie van de leeftijdsstructuur van een bevolking beschrijven, op dezelfde manier als Lotka en McKendrick hadden gedaan (zie de hoofdstukken 10 en 16). Met behulp van een continu-tijd model, noem

- $P(x, t)$ de populatie van leeftijd x op tijdstip t ;
- $m(x)$ de sterfte op leeftijd x ;
- $P_0(x)$ de leeftijdsstructuur van de bevolking op tijdstip $t = 0$;
- $b(t)$ de totale vruchtbaarheid van vrouwen op tijdstip t , d.w.z. het gemiddelde aantal kinderen dat een vrouw tijdens haar leven zou krijgen indien de leeftijdsspecifieke vruchtbaarheid zou blijven wat zij op tijdstip t is;
- f het aandeel van vrouwelijke geboorten;
- $h(x)$ de kansverdeling van de leeftijd van de moeder bij de geboorte van een kind ($\int_0^{+\infty} h(x) dx = 1$).

Met deze notaties en hypothesen kan de evolutie van de leeftijdsstructuur gemodelleerd worden door de partiële differentiaalvergelijking

$$\frac{\partial P}{\partial t}(x, t) + \frac{\partial P}{\partial x}(x, t) = -m(x)P(x, t),$$

met de beginvoorwaarde $P(x, 0) = P_0(x)$ en de randvoorwaarde

$$P(0, t) = b(t)f \int_0^{+\infty} h(x)P(x, t) dx,$$

waarbij $b(t)$ de parameter is die moet worden gecontroleerd. Als de totale vruchtbaarheid van vrouwen constant is en boven de kritische drempel

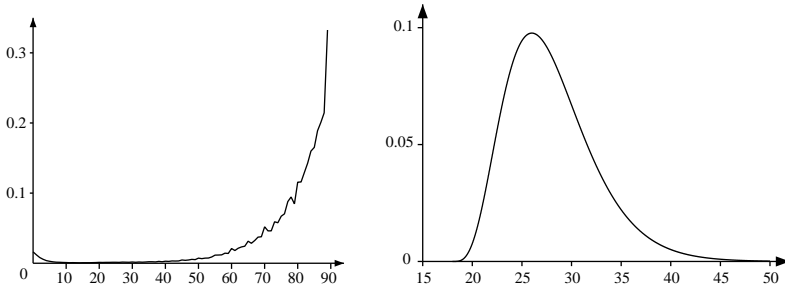
$$b^* = 1 / \left[f \int_0^{+\infty} h(x) e^{-\int_0^x m(y) dy} dx \right],$$

dan neemt de bevolking exponentieel toe. Dit criterium lijkt op dat van Lotka met de formule (10.2). Het team van Song Jian beschouwde ook de tijdsdiscrete versie van het model, die gelijkaardig is aan het model van Leslie (zie

hoofdstuk 21). Zij $P_{k,n}$ de bevolking van k jaar in jaar n . Stel op dezelfde wijze m_k , b_n en h_k voor. Dan is

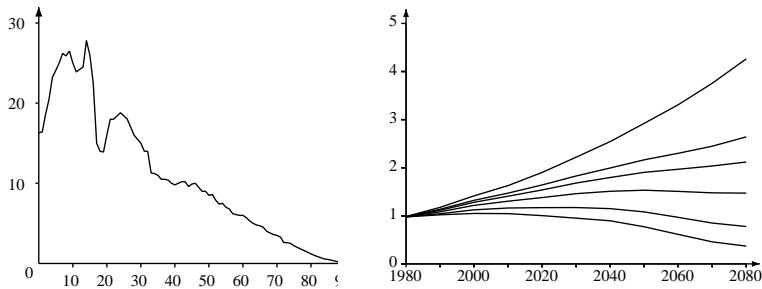
$$P_{k+1,n+1} = (1 - m_k)P_{k,n}, \quad P_{0,n+1} = b_n f \sum_{k \geq 0} h_k P_{k,n}.$$

Op basis van steekproevenquêtes kennen we het sterftecijfer m_k (figuur 25.2a), het aandeel van de vrouwelijke geboorten $f \approx 0,487$, de leeftijdsverdeling van de moeders h_k (figuur 25.2b), de beginvoorwaarde $P_{k,0}$ die de leeftijdsstructuur van de bevolking is in 1978 (figuur 25.3a) en het variëren van de totale vruchtbaarheid b (constant verondersteld tijdens elke simulatie), kon het team van Song Jian demografische projecties maken voor hun land met een tijdshorizon van honderd jaar, van 1980 tot 2080 (figuur 25.3b). Gezien de vereiste duizenden optellingen en vermenigvuldigingen (jaar n varieert tussen 0 en 100 jaar, leeftijd k tussen 0 en 90 jaar), was een computer noodzakelijk. In die tijd hadden in China slechts weinigen toegang tot dergelijke apparatuur, behalve degenen die voor het leger werkten. Song Jian, een vooraanstaand expert in raketgeleiding, was één van hen.



Figuur 25.2: (a) Sterfte (per jaar) als functie van de leeftijd in 1978. (b) Afgevlakte vorm van de vruchtbaarheid (per jaar) als functie van de leeftijd in 1978.

Uit de prognoses bleek dat zelfs indien China zijn vruchtbaarheidscijfer van 1978 van $b = 2,3$ kinderen per vrouw zou handhaven, wat net boven de kritische drempel ligt die op $b^* = 2,19$ wordt geschat, de bevolking zou toenemen van 980 miljoen in 1980 tot 2,12 miljard in 2080. Maar China gebruikte reeds bijna al het land dat voor landbouw kon dienen. Het had zelfs de neiging een deel van dit land te verliezen door woestijnvorming en verstedelijking. Hoe kan een dergelijke bevolking worden gevoed als de vooruitgang



Figuur 25.3: (a) Leeftijdspiramide in 1978. Horizontale as: leeftijd. Verticale as: bevolking (in miljoenen). (b) Demografische projecties (in miljarden) volgens verschillende hypothesen over het gemiddelde aantal kinderen per vrouw. Van onder naar boven: $b = 1,0$; $1,5$; $2,0$; $2,3$; $2,5$; $3,0$.

in landbouwopbrengsten niet voldoende is? Het is dezelfde vraag die Malthus twee eeuwen eerder had gesteld. Met een vruchtbaarheid van $b = 3,0$ in 1975 zou de bevolking in 2080 zelfs 4,26 miljard kunnen bereiken. Met $b = 2,0$ zou de bevolking rond het jaar 2050 een maximum van 1,53 miljard bereiken alvorens lichtjes te beginnen afnemen. Met $b = 1,5$ zou rond 2030 een maximum van 1,17 miljard worden bereikt. Met $b = 1,0$ zou het maximum slechts 1,05 miljard bedragen en rond 2000 worden bereikt. In die hypothese zou de bevolking pas tegen 2025 opnieuw het niveau van 1978 bereiken.

Het meest verrassende van dit werk waren de praktische gevolgen, in feite van ongeëvenaard belang in de geschiedenis van de mathematische populatiedynamica. Li Guangyuan toonde de resultaten van de simulaties van het team in december 1979 tijdens een symposium over bevolking in Chengdu, provincie Sichuan². In januari 1980 publiceerden Song Jian, Yu Jingyuan en Li Guangyuan deze resultaten in een Chinees economisch tijdschrift, waarbij zij overigens pleitten voor een één-kind-beleid. Zij stuurden hun artikel - Een rapport over kwantitatief onderzoek naar de kwestie van China's bevolkingsontwikkeling - naar China's topwetenschapper Qian Xuesen, die het met aanbeveling doorstuurde naar het hoofd van de administratie voor geboortepanning. De resultaten van Song Jian's team maakten een diepe indruk op de meeste politieke leiders. Deze waren reeds overtuigd van de noodzaak van een verhoogde geboortebeperving ondanks wat Marx had ge-

²Hier en hieronder vatten wij het gedetailleerde verslag van Susan Greenhalgh samen [1,2].

schreven (zie hoofdstuk 5) maar aarzelden nog over het niveau van de controle. In februari 1980 stelden de Staatsraad en het Centraal Comité van de Partij voor de Chinese bevolking een streefcijfer van 1,2 miljard vast voor het jaar 2000. In maart 1980 werden de resultaten van Song Jian's team gepubliceerd in het *Volksdagblad* (*Rénmín Ribào*). In april onderzocht een commissie van politieke leiders en bevolkingsdeskundigen de gevolgen van de bevolkingsgroei voor het milieu en de economie en concludeerde dat een één-kind-beleid noodzakelijk was om het door Deng Xiaoping gestelde doel voor het inkomen per hoofd van de bevolking in het jaar 2000 te bereiken. Het beleid werd officieel in september van datzelfde jaar en een open brief waarin het aan de bevolking werd uitgelegd, werd gepubliceerd op de eerste pagina van het *Volksdagblad*.

In 1983 zullen er nog steeds veel ongeoorloofde geboorten zijn. Besloten werd dat van elk echtpaar dat reeds twee kinderen had, een lid zou worden gesteriliseerd en dat elke verboden zwangerschap zou worden afgebroken. Vanaf 1984 mochten paren op het platteland met slechts één dochter echter toch een tweede kind krijgen. Het éénkindbeleid duurde tot het jaar 2015. De laatste jaren werden enkele aanpassingen doorgevoerd: als in een koppel zowel de man als de vrouw alleenstaande kinderen waren, dan mochten ze twee kinderen krijgen. De repressieve maatregelen tegen paren met meer dan één kind waren hard: overheidsambtenaren konden hun baan verliezen, er moest een dure boete worden betaald om de administratieve papieren te krijgen voor het onderwijs van een tweede kind, enzovoort. Kortom, het is moeilijk om in de geschiedenis van de wiskundige modellering nog een voorbeeld te vinden met zo'n sterke sociale impact. Natuurlijk was het werk van Song Jian en zijn medewerkers slechts een van de elementen die hebben geleid tot de keuze van het één-kind-beleid. Maar het lijkt een belangrijke rol te hebben gespeeld.

Zoals in de vorige hoofdstukken kan de rol van wiskundige modellering een punt van zorg zijn. Uitgaande van een reële situatie wordt een model gebouwd. Dit kan wiskundig worden geanalyseerd of gesimuleerd met een computer. Men kan dan begrijpen hoe het model zich gedraagt wanneer bepaalde parameters variëren. De wiskunde zegt echter niet of het model een getrouwe weergave van de werkelijkheid is. Sommige zeer belangrijke aspecten kunnen verwaarloosd zijn. Sommige modellen bevatten ook een objectieve functie, bijvoorbeeld de Chinese bevolking onder de 1,2 miljard houden. De wiskunde zegt niet of deze doelstelling adequaat was³.

³De bevolking werd in het jaar 2000 geschat op 1,264 miljard. Het inkomen per hoofd van de bevolking is tussen 1980 en 2000 ongeveer gestegen van 200 naar 1.000 dollar. Tegelijkertijd is de sekseverhouding extreem in het voordeel van jongens geworden, voornamelijk door sekseselectieve abortus.

In 1980 was Song Jian ook coauteur van de nieuwe editie van het boek *Technische Cybernetica* van Qian Xuesen, de „vader” van het Chinese ruimtevaartprogramma. Daarna bekleedde hij verschillende politieke functies op hoog niveau: vice-minister en hoofdwetenschapper-ingenieur van het Ministerie van Lucht- en Ruimtevaart (1981-1984), lid van het Centraal Comité van de Chinese Communistische Partij (1982-2002), voorzitter van de Staatscommissie voor Wetenschap en Technologie (1985-1998), Staatsraadslid (1986-1998) enzovoort. Hij publiceerde ook twee andere boeken die in het Engels zijn vertaald: *Bevolkingsbeheersing in China* (1985, met Tuan Chi-Hsien en Yu Jingyuan) en *Bevolking Systeem Controle* (1988, met Yu Jingyuan). Deze boeken ontwikkelen de theorie van optimale controle toegepast op populatiedynamica. Song Jian werd in 1991 verkozen tot lid van de Chinese Academie van Wetenschappen en in 1994 tot lid van de Academie van Ingenieurs, waarvan hij van 1998 tot 2002 voorzitter was.

Verdere lectuur

1. Greenhalgh, S.: Missile science, population science: The origins of China's one-child policy. *China Q.* 182, 253–276 (2005)
2. Greenhalgh, S.: *Just One Child, Science and Policy in Deng's China*. University of California Press (2008)
3. Meadows, D.H., Meadows, D.L., Randers, J., Behrens, W.W.: *The Limits to Growth, A Report for the Club of Rome's Project on the Predicament of Mankind*, 2nd edn. Universe Books, New York (1974)
4. Song, J.: Selected Works of J. Song. *Science Press*, Beijing (1999)
5. Song, J.: Some developments in mathematical demography and their application to the People's Republic of China. *Theor. Popul. Biol.* 22, 382–391 (1982)
6. Song, J., Yu, J.: *Population System Control*. Springer (1988)

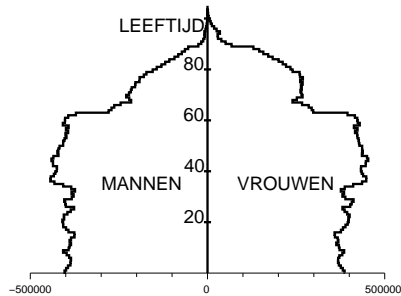
Hoofdstuk 26

Enkele hedendaagse problemen

In dit hoofdstuk wordt een kort overzicht gegeven van enkele hedendaagse problemen op het gebied van de mathematische populatiedynamica: de vergrijzing van de bevolking in de demografie; opkomende ziekten (AIDS, SARS, door vectoren overgebrachte ziekten) en het vaccinatiebeleid in de epidemiologie; het visserijbeleid in de ecologie; de verspreiding van genetisch gemodificeerde organismen in de populatiegenetica. De gespecialiseerde instellingen die in Frankrijk aan de modellering van deze problemen werken, worden vermeld. Ook wordt de nadruk gelegd op verschillende aspecten van het onderzoek.

In de demografie heeft zich de laatste decennia een betrekkelijk nieuw probleem aangediend: de vergrijzing van de bevolking. Dit probleem houdt niet alleen Frankrijk bezig (fig. 26.1), maar ook vele andere Europese landen en Japan. Vergrijzing heeft belangrijke economische en sociale gevolgen: pensioenstelsels, immigratiebeleid, enz. In Frankrijk worden door het Nationaal Instituut voor Demografische Studies (INED) en het Nationaal Instituut voor Statistische en Economische Studies (INSEE) mathematische modellen ontwikkeld om het vergrijzingsfenomeen te analyseren. Een van de moeilijkheden bij demografische prognoses is het feit dat de geboortecijfers in de loop van de tijd aanzienlijk kunnen variëren zonder dat dit zelfs maar tien jaar vooruit te voorzien is. Dit is met name opvallend wanneer men teruggaat naar de prognoses die in 1968 zijn gemaakt voor de Franse bevolking in 1985: deze prognoses hadden de daling van het geboortecijfer die zich in de jaren zeventig voordeed, niet voorzien. Het zou interessant zijn een overzicht te geven van alle op wiskundige modellen gebaseerde voorspellingen die onjuist bleken te zijn, met name die welke weerklank hebben gevonden in de media. Dit zou een tegenwicht vormen voor de indruk van „vooruitgang” die dit boek wekt, een indruk die de lezer wellicht al verdacht voorkwam na het lezen van het hoofdstuk over het Chinese één-kind-beleid. Wat dit laatste onderwerp betreft, is thans een nieuw probleem aan de orde: hoe kan het beleid worden afgezwakt om de in de komende decennia verwachte snelle vergrijzing te voorkomen. Ook hier dragen wiskundige modellen bij tot het debat.

In de epidemiologie is van de nieuwe problemen die zich in de laatste drie decennia wereldwijd hebben aangediend, de ontwikkeling van de AIDS-



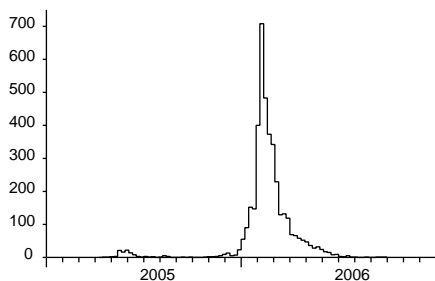
Figuur 26.1: Bevolkingspiramide van Frankrijk op 1 januari 2010. Bron: www.insee.fr.

epidemie in het bijzonder opvallend. Sommige modellen proberen de toekomst van de epidemie in recenter besmette landen zoals Rusland, India of China te voorspellen. Het is echter moeilijk te voorspellen of de epidemie zal vertragen zoals in West-Europa en Noord-Amerika of dat zij een belangrijk percentage van de bevolking zal bereiken zoals in sommige landen ten zuiden van de Sahara. Andere opkomende ziekten zoals Ebola in Afrika, West-Nijlkoorts in Noord-Amerika, SARS (*Severe Acute Respiratory Syndrome*), vogelgriep, chikungunya of H1N1-griep (Mexicaanse griep) zijn alle onder de loep genomen met wiskundige modellen, zij het met weinig succes.

Voor SARS was een moeilijkheid bij de modellering dat de epidemie relatief beperkt bleef binnen elk land, maar zich zeer snel van land tot land kon verspreiden (Hong Kong en China, Singapore, Canada...). Het willekeurige karakter van de epidemiecurves in elk nieuw brandpunt kon niet worden verwaarloosd. Zoals we in de hoofdstukken McKendrick en percolatie hebben gezien, zijn stochastische modellen meestal moeilijker te hanteren.

Voor de chikungunya-epidemie die tussen 2005 en 2006 op het eiland Réunion (een Frans overzees gebiedsdeel in de Indische Oceaan) heerste, zijn sommige modellen geïnspireerd op die van Ross voor malaria (zie hoofdstuk 12), aangezien beide ziekten door muggen worden overgebracht. Een belangrijk aspect waarmee rekening moest worden gehouden was de invloed van de seizoensgebondenheid. Tijdens de zuidelijke winter neemt de muggenpopulatie immers af, waardoor de ziekte minder snel wordt overgedragen. Dit blijkt uit fig. 26.2, waarin het aantal nieuwe gevallen is aangegeven dat wekelijks werd gemeld door een klein netwerk van een dertigtal huisartsen dat slechts een fractie van de bevolking van het eiland bestreekt. Het netwerk heeft gedurende verscheidene weken in september en oktober 2005 geen nieuwe gevallen geconstateerd, maar de overdracht van de ziekte ging nog

steeds door. Wiskundige modellen van de epidemie werden ontwikkeld aan het Nationale Instituut voor Gezondheids- en Medisch Onderzoek (INSERM) en aan het Instituut voor Tropisch Onderzoek (IRD). Ondanks deze modellen kon niemand voorspellen dat de epidemie niet zou uitsterven voor het einde van de zuidelijke winter van 2005, toen zij nog slechts enkele duizenden mensen had besmet. Uiteindelijk is bijna een derde van de bevolking van het eiland besmet geraakt, dat wil zeggen ongeveer 266.000 mensen. Hieruit blijkt, voor zo ver nog nodig, dat het voorspellen van de toekomst van epidemieën vrij moeilijk kan zijn en dat het niet zo gemakkelijk is om in de eerste dagen van een epidemie te onderscheiden of het een kleine of een grote epidemie zal worden. Er kan een parallel worden getrokken met weersvoorspellingen. Dit soort voorspellingen berust tegenwoordig op intensieve computersimulaties van ingewikkelde wiskundige modellen van de oceaan en de atmosfeer. Voorspellingen van meer dan een paar dagen zijn echter niet betrouwbaar.



Figuur 26.2: De chikungunya-epidemie op het eiland Réunion in 2005–2006. Aantal nieuwe gevallen per week gemeld door een klein netwerk van artsen als functie van de tijd. De eerste kleine piek werd in mei 2005 bereikt, de tweede grote piek in februari 2006. De getallen in deze figuur moeten met ongeveer 67 worden vermenigvuldigd om de werkelijke omvang van de epidemie te krijgen. Bron: www.invs.sante.fr.

Vanuit een meer theoretisch oogpunt heeft de chikungunya-epidemie de vraag doen rijzen hoe het begrip basisreproductiegetal \mathcal{R}_0 moet worden aangepast in modellen die veronderstellen dat de omgeving seizoensgebonden (bv. periodieke) schommelingen vertoont. Zo'n aanpassing is niet zo eenvoudig en dit geeft aanleiding tot enige bezorgdheid over de manier waarop de parameter \mathcal{R}_0 is gebruikt voor andere epidemieën die door seizoensinvloeden worden beïnvloed, zoals de H1N1-griep пандemie van 2009.

Een ander steeds zorgwekkender probleem dat modelleerders hebben trachten te analyseren is dat van de resistentie tegen geneesmiddelen (antibiotica,

antimalariamiddelen). Nog steeds in de epidemiologie is sinds de tijd van Daniel Bernoulli en d'Alembert de terugkerende vraag hoe de kosten en baten tegen elkaar moeten worden afgewogen wanneer de injectie van een vaccin een potentieel risico inhoudt nog altijd omstreden en dat zal wellicht zo blijven naarmate de gevoeligheid voor risico's verandert. Zo heeft het Franse ministerie van Volksgezondheid, naar aanleiding van bepaalde suggesties dat het vaccin tegen hepatitis B in een zeer klein aantal gevallen ernstige complicaties zou kunnen veroorzaken, in 1998 de vaccinatiecampagne op scholen stopgezet, ook al leek het risico te verwaarlozen in vergelijking met het risico te overlijden na besmetting met het hepatitis B-virus.

In de ecologie levert de studie van de dynamiek van vispopulaties nog veel problemen op. Toch wordt zij geacht te dienen als wetenschappelijke basis voor de keuze van vangstquota en andere beperkingen. De overbevissing van de ansjovis in de Golf van Biskaje en van de rode tonijn in de Middellandse Zee zijn slechts twee recente voorbeelden. Aangezien de raming van het visbestand vaak onbetrouwbaar is, moeten modellen die van dergelijke gegevens gebruik maken, met de nodige omzichtigheid worden gehanteerd. In Frankrijk wordt dit soort onderzoek hoofdzakelijk verricht door het Instituut voor Onderzoek van de Zee (IFREMER). Sommige wiskundige modellen hebben ook een rol gespeeld in vroegere besluiten van de Internationale Walvisvangstcommissie.

In de populatiegenetica is de verspreiding van genetisch gemodificeerde organismen eveneens een controversieel onderwerp dat sommige onderzoekers hebben geprobeerd te bestuderen met behulp van „reactie-diffusie” modellen die zijn geïnspireerd op die van Fisher (zie hoofdstuk 20). Dit is het gebied van het Nationaal Instituut voor Onderzoek in Agronomie (INRAE).

Wat de meer theoretische kant van het onderzoek betreft, kunnen worden vermeld:

- de werken over partiële differentiaalvergelijkingen zoals diffusievergelijkingen (zie hoofdstuk 20) of leeftijdsgestructureerde vergelijkingen (zie hoofdstuk 16);
- de werken over stochastische modellen met of zonder ruimtelijke dimensie (zie de hoofdstukken 16 en 22), met inbegrip van die over willekeurige netwerken die de verspreiding van epidemieën modelleren en die welke deterministische benaderingen zoeken.

Dit type onderzoek wordt voornamelijk uitgevoerd door toegepaste wiskundigen. De laatste jaren zijn aan de Franse universiteiten en andere instellingen voor hoger onderwijs verschillende masteropleidingen in mathematische biologie ingevoerd.

Net als andere wetenschapsgebieden wordt de wiskundige studie van de populatiedynamica hoofdzakelijk georganiseerd door:

- „geleerde genootschappen”: Nederlandse Vereniging voor Theoretische Biologie (sinds 1970), Vereniging voor Mathematische Biologie (1973), Franstalige Vereniging voor Theoretische Biologie (1985), Japanse Vereniging voor Mathematische Biologie (1989), Europese Vereniging voor Mathematische en Theoretische Biologie (1991) enz.
- gespecialiseerde tijdschriften: *Acta Biotheoretica* (sinds 1935), *Bulletin of Mathematical Biology* (1939), *Mathematical Biosciences* (1967), *Journal of Mathematical Biology* (1974), *Mathematical Medicine and Biology* (1984), *Mathematical Population Studies* (1988), *Mathematical Biosciences and Engineering* (2004), *International Journal of Biomathematics* (2008), *Biomath* (2012) enz.
- conferenties (jaarlijkse bijeenkomst van de Vereniging voor Mathematische Biologie, Computatieve en wiskundige populatiedynamica, Europese Conferentie over Wiskundige en Theoretische Biologie enz.).

Er is alleen verwezen naar de elementen die expliciet beweren zich op het raakvlak te bevinden tussen wiskunde en de toepassingen ervan op de populatiedynamica. Maar voor elk specifiek gebied (demografie, ecologie, populatiegenetica, epidemiologie, enz.) kan men soortgelijke elementen vinden met een variabele dosis wiskundige modellering.

Tot slot wordt de geïnteresseerde lezer verzocht een kijkje te nemen in de originele artikelen die op het *World Wide Web* beschikbaar zijn. De adressen staan vermeld in de referenties aan het eind van elk hoofdstuk. Zoals Ronald Fisher ooit over Mendel schreef:

„De geschiedenis van de wetenschap heeft veel te lijden gehad van het gebruik door leraren van tweedehands materiaal en de daaruit voortvloeiende uitwissen van de omstandigheden en de intellectuele atmosfeer waarin de grote ontdekkingen van het verleden werden gedaan. Een studie uit de eerste hand is altijd leerzaam en vaak vol verrassingen.”

Verdere lectuur

1. Bacaër, N.: Approximation of the basic reproduction number \mathcal{R}_0 for vector-borne diseases with a periodic vector population. *Bull. Math. Biol.* 69, 1067–1091 (2007)
2. Levin, S.A.: Mathematics and biology, the interface. www.bio.vu.nl/nvtb/

Afbeeldingen

- Pag. 5. Portret van Thomas Murray (ca. 1687) in het bezit van de *Royal Society* in Londen. Chapman, S.: Edmond Halley, F.R.S. 1656–1742. *Notes Rec. R. Soc. Lond.* 12, 168–174 (1957) © The Royal Society.
- Pag. 11. Portret van Emanuel Handmann (1753) in het bezit van het Kunstmuseum in Basel. *Leonhard Euler 1707–1783, Beiträge zu Leben und Werk.* Birkhäuser, Basel (1983)
- Pag. 17. Portret dat zich ooit in de Sint-Pieterskerk bevond en waarschijnlijk verloren is gegaan tijdens de Slag om Berlijn in 1945. Reimer, K.F.: Johann Peter Süßmilch, seine Abstammung und Biographie. *Arch. soz. Hyg. Demogr.* 7, 20–28 (1932)
- Pag. 23. Portret van Johann Niclaus Grooth (ca. 1750–1755) in het bezit van het Natuurhistorisch Museum in Basel. Speiser, D.: *Die Werke von Daniel Bernoulli*, Band 2. Birkhäuser, Basel (1982)
- Pag. 30. Portret van Maurice Quentin de La Tour (1753) in het bezit van het Louvre in Parijs.
- Pag. 33. Portret van John Linnell (1833) in het bezit van *Haileybury College*, in Engeland. Habakkuk, H.J.: Robert Malthus, F.R.S. (1766-1834). *Notes Rec. R. Soc. Lond.* 14, 99–108 (1959)
- Pag. 37. Gravure van Flameng (1850). Quetelet, A.: Pierre-François Verhulst. *Annu. Acad. R. Sci. Lett. B.-Arts Belg.* 16, 97–124 (1850)
- Pag. 42. Heyde, C.C., Seneta, E.: *I. J. Bienaymé, Statistical Theory Anticipated.* Springer-Verlag, New York (1977) © Académie des sciences, Institut de France.
- Pag. 46. Bateson, W.: *Mendel's Principles of Heredity.* Cambridge University Press (1913)
- Pag. 51. Pearson, K.: *The Life, Letters, and Labors of Francis Galton*, vol. 1. Cambridge University Press (1914)
- Pag. 51. Portret van Watson in de bibliotheek van *Trinity College*, Universiteit van Cambridge. Kendall, D.G.: Branching processes since 1873. *J. Lond. Math. Soc.* 41, 385–406 (1966)
- Pag. 57. Archief Alfred J. Lotka. © Princeton University Library.
- Pag. 61. Titchmarsh, E. C.: Godfrey Harold Hardy 1877–1947. *Obit. Not. Fellows R. Soc.* 6, 446–461 (1949)
- Pag. 64. Stern, C.: Wilhelm Weinberg. *Genetics* 47, 1–5 (1962)
- Pag. 67. G.H.F.N.: Sir Ronald Ross 1857–1932. *Obit. Not. Fellows R. Soc.* 1, 108–115 (1933) © The Royal Society.

- Pag. 76. Whittaker, E.T.: Vito Volterra 1860–1940. *Obit. Not. Fellows R. Soc.* 3, 690–729 (1941)
- Pag. 80. Yates, F., Mather, K.: Ronald Aylmer Fisher, 1890–1962. *Biog. Mem. Fellows R. Soc.* 9, 91–120 (1963) © The Royal Society/Godfrey Argent Studio.
- Pag. 84. Yates, F.: George Udny Yule. *Obit. Not. Fellows R. Soc.* 8, 308–323 (1952)
- Pag. 92. Heyde, C.C., Seneta, E. (eds.): *Statisticians of the Centuries*. Springer (2001)
- Pag. 102. britannica.com/EBchecked/topic/252257/J-B-S-Haldane © Bassano and Vandyk Studios.
- Pag. 109. Hill, W.G.: Sewall Wright, 21 December 1889–3 March 1988. *Biog. Mem. Fellows R. Soc.* 36, 568–579 (1990) © Llewellyn Studios, Chicago.
- Pag. 106. Nybølle, H.C.: Agner Krarup Erlang f. 1. Januar 1878 - d. 3. Februar 1929. *Mat. Tidsskr. B*, 32–36 (1929)
- Pag. 118. Tikhomirov, V.M.: A.N. Kolmogorov. In: Zdravkovska, S., Duren, P.L. (eds.) *Golden Years of Moscow Mathematics*, 2nd edn., Pag. 101–128. American Mathematical Society (2007)
- Pag. 118. *I. G. Petrowsky Selected Works Part I*. Gordon and Breach, Amsterdam (1996) © Taylor and Francis Books UK.
- Pag. 123. Foto door Denys Kempson. Crowcroft, P.: *Elton's Ecologists, a History of the Bureau of Animal Population*. University of Chicago Press (1991)
- Pag. 126. © Geoffrey Grimmett.
- Pag. 132. Charlesworth, B., Harvey, P.: John Maynard Smith, 6 January 1920–19 April 2004. *Biog. Mem. Fellows R. Soc.* 51, 253–265 (2005) © The Royal Society.
- Pag. 138. © Samuel Schläefli / ETH Zürich.
- Pag. 145. Selected works of J. Song. *Science Press*, Beijing (1999) © Song Jian.

Inhoudsopgave

Inleiding	iii
1 De rij van Fibonacci (1202)	1
2 Halley's sterftetabel (1693)	4
3 Euler en de geometrische groei van populaties (1748–1761)	10
4 Daniel Bernoulli, d'Alembert en de inenting van pokken (1760)	22
5 Malthus en de belemmeringen voor geometrische groei (1798)	33
6 Verhulst en de logistische vergelijking (1838)	37
7 Bienaymé en het uitsterven van familienamen (1845)	42
8 Mendel en erfelijkheid (1865)	46
9 Galton, Watson en het probleem van uitsterven (1873–1875)	50
10 Lotka en de stabiele bevolkingstheorie (1907–1911)	57
11 De wet van Hardy-Weinberg (1908)	61
12 Ross en malaria (1911)	66
13 Lotka, Volterra en het predator-prooi systeem (1920–1926)	72
14 Fisher en natuurlijke selectie (1922)	79
15 Yule en evolutie (1924)	83
16 McKendrick over epidemische modellering (1926–1927)	91
17 Haldane en mutaties (1927)	101
18 Erlang over het uitstervingsprobleem (1929)	105
19 Wright en willekeurige genetische drift (1931)	109
20 De verspreiding van genen (1937)	115
21 De Leslie-matrix (1945)	122
22 Percolatie en epidemieën (1957)	126
23 Speltheorie en evolutie (1973)	132
24 Chaotische populaties (1974)	138
25 China's één-kind politiek (1980)	145
26 Enkele hedendaagse problemen	152

Dit boek schetst de geschiedenis van de populatiedynamica - een onderzoeksveld dat nauw verbonden is met genetica, ecologie, epidemiologie en demografie, waarin de wiskunde belangrijke inzichten heeft gebracht. Het geeft een overzicht van de ontstaansgeschiedenis van verschillende belangrijke concepten, zoals exponentiële groei, van Euler en Malthus tot het Chinese één-kind-beleid, de ontwikkeling van stochastische modellen, van de wetten van Mendel en het vraagstuk van het uitsterven van familienamen tot de percolatietheorie voor de verspreiding van epidemieën, en chaotische dynamica, waarin determinisme en toeval samengaan.

Met de recente vooruitgang op het gebied van automatische vertaling is het monopolie van één enkele taal in de wetenschappelijke literatuur niet langer gerechtvaardigd. De groeiende taalvervreemding aan universiteiten kan worden gestopt. Met deze zorgvuldig herziene Nederlandse vertaling hopen wij daaraan bij te dragen.

ISBN : 979-10-343-6356-8



15€