

Nicolas Bacaër
med
Carl-Joar Karlsson
och med
Torsten Lindström, Philip Gerlee,
Torbjörn Lundh, Peter Olofsson

Axplock
i den matematiska
populationsdynamikens
historia



Axplock i den matematiska populationsdynamikens historia

Nicolas Bacaër

med

Carl-Joar Karlsson

och med

Torsten Lindström, Philip Gerlee,

Torbjörn Lundh, Peter Olofsson

Nicolas Bacaër
Institut de recherche pour le développement
nicolas.bacaer@ird.fr

Carl-Joar Karlsson, Philip Gerlee, Torbjörn Lundh
Göteborgs universitet och Chalmers tekniska högskola
carljoar@chalmers.se, gerlee@chalmers.se, torbjorn.lundh@chalmers.se

Torsten Lindström
Linnéuniversitetet
Torsten.Lindstrom@lnu.se

Peter Olofsson
Högskolan i Jönköping
Peter.Olofsson@ju.se

Läsare som vill köpa pappersversionen av denna bok kan skicka ett e-postmeddelande till nicolas.bacaer@ird.fr.

Omslagsbild: Svenska och norska paviljongen på världsutställningen i Paris 1878 (nu i Courbevoie).

Titre original: Histoires de mathématiques et de populations
© Cassini, Paris, 2008

Pour l'édition suédoise:
© Nicolas Bacaër, Paris, 2022
ISBN: 979-10-343-9008-3
Dépôt légal: juin 2022

Inledning

Populationsdynamik är en gren inom vetenskapen där man tar sig an uppgiften att förklara mekanismerna bakom hur storlek och sammansättning varierar med tiden hos biologiska populationer, exempelvis hos människor, djur, växter eller mikroorganismer. Området påminner om, men skiljer sig samtidigt helt och hållet från, det mer beskrivande området befolkningsstatistik. Båda områdena delar emellertid matematiken som gemensamt språk.

De tvärvetenskapliga dragen i populationsdynamiken blir tydligast i ljuset av att den sträcker sig över många områden, bland annat matematik, demografi, genetik, ekologi och epidemiologi. Den presenteras på gott och ont sällan som en helhet trots att dess tillämpningsområden delar många gemensamma drag. Ett nämnvärt undantag är Alain Hillions bok *Matematisk populationsdynamik*¹ på franska. Här presenteras ämnet ur ett matematiskt perspektiv och man skiljer därmed på modeller i diskret tid (tiden antar heltalsvärden), kontinuerlig tid (tiden antar reella värden), deterministiska (framtiden är känd om nuläget är känt) och stokastiska modeller (framtiden kan bara förutsägas i termer av sannolikheter). Boken behandlar sedan diskreta deterministiska modeller, kontinuerliga deterministiska modeller, diskreta stokastiska modeller och kontinuerliga stokastiska modeller i en matematisk-logisk ordning.

I den här boken diskuterar jag samma ämne ur ett historiskt perspektiv. Forskning förklaras i sitt sammanhang. Korta biografier över forskarna ingår i framställningen. Min förhoppning är att detta skall göra ämnet lättare att förstå för läsare utan matematisk bakgrund emedan problemens ursprung klargörs. Orsaken till att jag skrev boken var dock inte bara vetenskapshistorisk. Jag ville att detta verk även skulle fungera som en inledning till matematisk modellering och jag ville att detaljerna i beräkningarna skulle bli tydliga så att läsaren kunde förstå modellernas begränsningar. Jag placerade samtidigt de mer tekniska detaljerna i separata rutor så dessa kan skummas vid en första genomläsning. Det avslutande kapitlet tar upp en rad populationsdynamiska problem från vår samtid där lösningarna förväntas kräva matematisk analys. Källförteckningen i slutet av varje kapitel innehåller även hänvisningar till sidor på nätet där originalartiklar finns tillgängliga.

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

Denna bok har inte som mål att i detalj beskriva alla bidrag som har haft betydelse för den matematiska populationsdynamikens utveckling under de senaste seklen. Det urval som jag har gjort omfattar naturligtvis ett visst mått av godtycke och det gäller framför allt de senaste decenniernas bidrag. Min förhoppning är ändå att mitt urval är så representativt att personer som är aktiva inom området inte känner sig stötta.

Jag hoppas att det jag skrivit kan verka intresseväckande bland följande målgrupper:

- Gymnasie- och universitetsstudenter som är nyfikna på kopplingar mellan obligatoriska matematikkurser och matematiska problem i deras omvärld alternativt studenter som står i beråd att skriva ett självständigt arbete med kopplingar till populationsdynamik.
- Matematiklärare på olika nivåer som arbetar med att göra sina kurser mer intressanta. Kunskaper om det fyra räkneseätten är tillräckliga för att förstå stora delar av kapitel 1, 2 och 5. Kapitel 3 kan användas för att beskriva tillämpningar av logaritmer. Boken behandlar rekursionsformler i kapitel 1, 3, 8, 11, 14, 21, 23, 24; ordinära differentialekvationer i kapitlen 4, 6, 12, 13, 16; partiella differentialekvationer i kapitlen 20, 25; en integralekvation i kapitel 10; och tillämpningar av sannolikhetslära i kapitlen 2, 7, 8, 9, 15, 16, 17, 18, 19, 22.
- Personer som är bekanta med demografi, epidemiologi, genetik eller ekologi och som vill jämföra sitt område med andra områden där likartade matematiska resultat används.
- Läsare som är intresserade av vetenskapshistoria.

Carl-Joar Karlsson (15 kapitel), Torsten Lindström (6 kapitel), Philip Gerlee (4 kapitel), Torbjörn Lundh (1 kapitel) och Peter Olofsson (1 kapitel) har läst igenom och lämnat förslag på korrektur till Maskinöversättningen av DeepL.

Kapitel 1

Fibonaccis talföljd (1202)

År 1202 publicerade Leonardo av Pisa, även kallad Fibonacci, en bok för att göra det indiska decimalsystemet tillgängligt i Europa. Det var redan vid här laget vedertaget i arabvärlden. I boken finns många exempel och ett av dem beskriver tillväxten av en kaninpopulation. Det är ett av de äldsta kända exemplen på en populationsdynamisk matematisk modell.

Leonardo av Pisa, som fick namnet Fibonacci långt efter sin död, föddes omkring 1170 i republiken Pisa. Pisa stod då på toppen av sin kommersiella och militära makt i Medelhavssfären. På uppdrag av republiken sändes Fibonaccis far omkring 1192 ut till hamnen i Bejaia (i dagens Algeriet) som föreståndare för en av dess handelsplatser. Hans son anslöt sig kort därefter till honom och förberedde sig för att bli handelsman. Leonardo började lära sig det decimala talsystem som araberna hade fört med sig från Indien. Det används fortfarande används i nästan samma form än i dag med sina siffror: 0, 1, 2, 3, 4, 5, 6, 7, 8 och 9. Under sina affärsresor runt Medelhavet jämförde han olika talsystem och studerade arabisk matematik. Tillbaka i Pisa skrev han 1202 klart en bok på latin med titeln *Liber abaci* ("Räkneboken") där han förklarade det nya talsystemet och visade hur man använde det för bokföring, viktberäkningar, valutaväxlingar, räntesatser och många andra tillämpningar. I den här boken samlade han även många bland araberna redan välkända resultat inom algebra och aritmetik.

Fibonacci beskrev i sin bok något som man idag skulle klassa som ett populationsdynamiskt problem. Men det framställdes bara som en räkneövning bland många andra ämnen utan anknytning till biologin eller med några krav på logisk ordning. Det föregående avsnittet i boken handlar om perfekta tal som är summan av sina faktorer, exempelvis $28 = 14 + 7 + 4 + 2 + 1$. Nästa avsnitt handlar om fördelningen av pengar mellan fyra personer. Problemet är ekvivalent med ett linjärt system bestående av fyra ekvationer. En översättning från latin av det populationsdynamiska problemet lyder:

"En man hade ett par kaniner i ett hägn. Hur många kaniner kan avlas fram från det ursprungliga paret under ett år om varje kaninpar föder ett nytt par varje månad och om kaninerna blir könsmogna först efter den andra månaden?"

Antag att det finns ett par nyfödda kaniner i början av den första månaden. Detta par kommer inte att vara köns mogna efter en månad och det kommer fortfarande att finnas bara ett par kaniner i början av den andra månaden. Detta kaninpar kommer att föda ytterligare ett par i början av den tredje månaden. Det finns alltså två par totalt. Det första kaninparet kan föda ett nytt par i början av den fjärde månaden. Men det andra kaninparet kommer ännu inte att vara köns moget. Det kommer bara att finnas tre par kaniner.

Låt P_n vara antalet kaninpar i början av månad n . Antalet kaninpar P_{n+1} i månad $n + 1$ är summan av antalet par P_n i månad n och antalet nyfödda par i månad $n + 1$. Men endast de kaninpar som är minst två månader gamla föder nya kaninpar i månad $n + 1$. Dessa är de par som fanns redan i månaden $n - 1$ och deras antal är P_{n-1} . Därmed är

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

Detta är en rekursionsformel som beskriver ett samband mellan populationsstorleken månad $n + 1$ och populationsstorleken under de två föregående månaderna. Fibonacci kunde med hjälp av detta bygga en tabell, där $1 + 1 = 2$, $1 + 2 = 3$, $2 + 3 = 5$, $3 + 5 = 8$ osv.

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

De facto betraktade Fibonacci situationen i månad $n = 2$ som utgångsläge. Eftersom $P_{14} = 144 + 233 = 377$ fick han slutligen 377 par kaniner tolv månader senare. Han lade märke till att denna sifferföljd kunde fortsättas i oändlighet.

Efter 1202 skrev Fibonacci flera andra böcker, till exempel *Practica geometriæ* ("Praktisk geometri") 1220 och *Liber quadratorum* ("Lärobok om kvadrater") 1225. Hans rykte ledde till ett möte med kejsar Fredrik II, som uppskattade vetenskap. År 1240 beviljade republiken Pisa Fibonacci en årlig pension. Hans dödsår är okänt.

Fibonaccis kaninproblem förblev bortglömt under de påföljande århundradena och fick under den här tiden inget inflytande på utvecklingen av populationsdynamiska modeller. Flera forskare mötte samma talföljd i sina studier men hänvisade inte till Fibonacci eller till något populationsproblem. Flera av Keplers böcker innehåller en anmärkning om att P_{n+1}/P_n konvergerar mot det gyllene snittet $\phi = (1 + \sqrt{5})/2$ när n går mot oändligheten. Många populationsmodeller uppvisar exponentiell tillväxt (se kapitlen 3 och 21) och

detta resultat är ett specialfall av denna egenskap. År 1728 härledde Daniel Bernoulli den exakta formeln

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

i samband med studie av talföljder i allmänhet. Fibonaccis fullständiga verk publicerades under 1800-talet. Från och med den här tidpunkten nämns kopplingen mellan talföljden (P_n) och Fibonaccis arbete ofta i matematisk litteratur.

De antaganden som leder till Fibonaccis talföljd i kaninpopulations-exemplet är emellertid orealistiska. I modellen finns ingen ansats till att ta med vare sig mortalitet eller könsindelning. Talföljden har dock under de senaste decennierna dragit till sig annat biologiskt intresse. Flera växter innehåller strukturer som är kopplade till talen P_n , bland annat 8 och 13 i tallkottar och 34 och 55 i solrosor. En vetenskaplig tidskrift, *The Fibonacci Quarterly*, är även helt dedicerad för Fibonacci-talföljdens egenskaper och tillämpningar!

Ytterligare läsning

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)

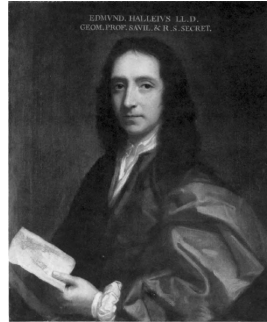
Kapitel 2

Återstående livslängd vid given ålder enligt Halley (1693)

År 1693 började den engelske astronomen Edmond Halley intressera sig för staden Breslaus (Wrocław i dagens Polen) folkbokföring, som hade överlåtits till *Royal Society* (det kungliga vetenskapssamfundet) av Caspar Neumann. Han kunde med hjälp av den beräkna hur många personer som överlevde till olika åldrar i varje årskull. Uppgifterna kunde användas för att beräkna priser på pensioner och livräntor. Vi kommer att koppla detta arbete till Halleys liv och den tidiga utvecklingen av "politisk aritmetik" och sannolikhetsteori här. Dessa områden intresserade även personer som Graunt, Petty, De Witt, Hudde, Huygens, Leibniz och de Moivre.

Edmond Halley föddes nära London 1656. Hans far var en rik tvåtillverkare. Edmond blev tidigt intresserad av astronomi. Han började studera vid *Queen's College* vid Oxfords universitet. När Greenwich-observatoriet invigdes 1675 fick Halley besöka John Flamsteed, den dåvarande innehavaren av tjänsten som *Astronomer Royal* (kunglig astronom). Han avbröt sina studier mellan 1676 och 1678 för att åka till ön Sankt Helena för att upprätta en förteckning över de stjärnor som kan observeras från det södra halvklotet. Han blev han medlem i *Royal Society* vid återkomsten till England. Han publicerade de observationer som han hade gjort om vindriktningar under sin resa till Sankt Helena. År 1684 besökte han Newton i Cambridge för att diskutera sambandet mellan Keplers lagar för planetrörelse och den attraktionskraft som solen utövar. Han uppmuntrade Newton att skriva *Naturvetenskapens matematiska principer* och finansierade sedermera utgivningen av boken själv. Under den här tiden arbetade han som sekreterare i *Royal Society*. År 1689 konstruerade han en dykarklocka, som han själv använde.

En teolog vid namn Caspar Neumann samlade ungefär samtidigt in folkbokföringsuppgifter i Breslau (Wrocław i dagens Polen). Staden var vid den här tidpunkten en del av det habsburgska riket. Uppgifterna omfattade även den ålder vid vilken människor hade dött. Man kunde alltså använda



Figur 2.1:
Halley (1656–1742)

uppgifterna för att upprätta en tabell som angav sannolikheten för att leva fram till en viss ålder.

En motsvarande tabell publicerades för första gången i London 1662 i en bok med titeln *Naturliga och politiska observationer grundade på löpsedlar om dödlighet*. Denna bok brukar anses vara den grundläggande texten för både statistik och demografi och den omgivs fortfarande av en viss mystik. Man vet inte om den skrevs av John Graunt, den handelsman i London som anges på bokomslaget, eller om den skrevs av hans vän William Petty, en av grundarna av *Royal Society*. I bokens tabell försökte man dra nytta av nyhetsblad som innehöll uppgifter om begravningar och dop i London sedan början av 1600-talet. Nyhetsbladen användes i första hand för att informera folket om de återkommande farsotsepidemierna. Dödsorsaken sågs som viktigare än den avlidnes ålder. För att få fram en tabell som angav chansen att överleva en tid efter en viss ålder så måste Graunt och Petty gissa hur dödsorsakerna hängde ihop med olika åldersgrupper. Gissningarna behäftade deras tabell med många fel. Boken blev ändå tillräckligt framgångsrik för att ges ut i fem upplagor mellan 1662 och 1676. Flera Europeiska städer började nu efterhand offentliggöra dödsorsaker i nyhetsblad på samma sätt som i London.

Nästan trettio år efter detta skickade Neumann på förslag av Leibniz sina demografiska data från staden Breslau för åren 1687–1691 till Henry Justel, dåvarande sekreterare på *Royal Society*. Justel dog kort därefter. Halley fick ändå tag på uppgifterna, analyserade dem och offentliggjorde 1693 sina slutsatser i tidskriften *Philosophical Transactions of the Royal Society*. Rubriken på hans uppsats lyder ungefär *En uppskattning av mortalitetet hos människor utifrån märkliga tabeller över födslo- och begravningsdata i Breslau och ett försök att fastställa priset på livräntor*.

Halley noterade att antalet födda i Breslau var i stora drag lika stort

som antalet döda under en femårsperiod. Befolkningens storlek var därmed konstant. Analysen kunde således förenklas om man antog att befolkningen befann sig i jämviktsläge. Vi antar alltså att antalet födslar (kalla det P_0), befolkningens storlek, åldersklassen k (P_k) och det årliga antalet dödsfall i åldern k (D_k) är konstanta över tid. Denna egenskap hos uppgifterna från Breslau är mycket intressant. Den snabba urbaniseringen gjorde en sådan förenkling omöjlig i en snabbt växande storstad som London.

Tabell 2.1: Halleys tabell visar antalet överlevande i P_k i ålderklass k bland 1238 nyfödda.

k	P_k	k	P_k	k	P_k	k	P_k	k	P_k	k	P_k
1	1000	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

Uppgifterna från Breslau hade ett medelvärde på 1238 födda per år: Detta är det värde som Halley tog för P_0 . I princip kunde han också utifrån uppgifterna beräkna det årliga medelvärdet D_k av antalet dödsfall bland människor i åldern k för alla $k \geq 0$. Med hjälp av formeln

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

kunde han konstruera en tabell 2.1 som ger P_k . Omvänt kan man hitta de värden på D_k som han använde utifrån formeln $D_k = P_k - P_{k+1}$: $D_0 = 238$, $D_1 = 145$, $D_2 = 57$, $D_3 = 38$ och så vidare. I själva verket organiserade Halley om sina resultat en aning, antingen för att få närmvärden (exempelvis D_1 , som har ändrats något så att $P_1 = 1000$) eller för att jämma ut vissa oregelbundenheter på grund av få dödsfall i hög ålder i en femårig studie. Genom att ta summan av alla siffror P_k i tabellen fick Halley en uppskattning av den totala befolkningen i Breslau: ungefär 34 000 personer¹. Denna metod

¹För personer över 84 år nämnde Halley bara att deras totala antal uppgick till 107.

har sammantaget fördelen att den inte krävde någon allmän folkräkning utan endast kunskap om antalet födselar och dödsfall med dödsåldrar under ett fåtal år.

Halleys uppgifter användes i många olika arbeten på 1700-talet (se kapitel 4). Även om värdena för P_k var specifika för staden Breslau kunde man faktiskt betrakta förhållandet P_{k+1}/P_k som sannolikheten att överleva till ålder $k + 1$ betingat att man redan uppnått ålder k . Denna sannolikhet kunde rimligen även användas för befolkningen i andra europeiska städer vid den här tiden. Exempelvis förväntar vi oss utifrån tabellen att ett ettårigt barn har en chans på 661/1000 att nå 10 års ålder samtidigt som det har en chans på 598/1000 att nå 20 års ålder.

Halley använde sin tabell för att beräkna priset på livräntor. Under 1500- och 1600-talen sålde flera städer och stater livränteåtaganden till sina medborgare för att balansera sin ekonomi. Köparna fick oberoende av ålder varje år fram till sin död en fast summa pengar som motsvarade en andel av den summa som ursprungligen betalats. Andelen sattes vanligtvis till dubbel ränta vid ifrågavarande tidpunkt. Förvaltaren riskerade naturligtvis att gå i konkurs om alltför många personer med lång kvarvarande livslängd satsade på sådana livräntor. Någon möjlighet att hantera detta problem utan tillförlitliga överlevnadsuppgifter förelåg inte.

Redan 1671 hade Johan De Witt, Nederländernas premiärminister, och Johannes Hudde, en av borgmästarna i staden Amsterdam, tagit sig an problemet med att beräkna priset på livräntor. Eftersom de fruktade en invasion av franska trupper behövde de samla in pengar för sin armé. De hade uppgifter om personer som hade köpt livräntor för flera årtionden tidigare, speciellt om den ålder när livräntorna hade köpts och den ålder då personerna hade dött. De hade lyckats beräkna priset på livräntor mer eller mindre korrekt, men deras metod glömdes därefter bort. Holland invaderades året därpå och De Witt lynchades av folkmassan.

Halley tog sig an problemet på nytt år 1693 med hjälp av tabellen från Breslau och med en ränta på 6%. Beräkningsmetoden var enkel. Låt i vara räntesatsen. Låt R_k vara det pris till vilket en person i åldern k kan köpa en livränta på, säg ett pund per år. Denna person har en sannolikhet P_{k+n}/P_k att fortfarande leva vid ålder $k + n$. Det pund som staten lovar att betala om han når denna ålder kan erhållas genom att placera $1/(1+i)^n$ pund av den ursprungliga summan till räntesatsen i . Om man gör det förenklade antagandet att den ursprungliga summan endast används för att betala livräntan, bör priset vara

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

Halley fick på detta sätt fram tabell 2.2, som visar faktorn R_k som den önskade livräntan måste multipliceras med för att få det nödvändiga priset. En 20-årig man skulle alltså varje år få $1/12,78 \approx 7,8\%$ av den ursprungliga summan. Men en man på 50 år skulle få $1/9,21 \approx 10,9\%$, eftersom den återstående levnadstiden är kortare. Observera att en dubbelt så hög ränta skulle motsvara en livränta som motsvarar 12% av det ursprungliga beloppet alternativt ett motsvarande pris på 8,33 gånger livräntan.

Tabell 2.2: Multiplikationsfaktorer vid prissättning av livräntor.

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

Dessa beräkningar är naturligtvis rätt tråkiga. Halley kunde dock använda logaritmtabeller för att snabbare få fram den allmänna termen $P_{k+n}/(1+i)^n$. Eftersom han inte visade värden för P_k över 84 år är det inte möjligt att kontrollera hans beräkningar exakt. Halleys arbete fick dock ingen omedelbar uppmärksamhet: Under årtionden fortsatte livräntor att säljas i England och på andra håll till ett pris som var oberoende av köparens ålder och kraftigt rabatterat, exempelvis 7 gånger livräntan.

Tabellerna intresserade många forskare redan under Halleys tid. Holländaren Christiaan Huygens, som redan 1657 hade publicerat ett häfte om sannolikhetsteori, diskuterade 1669 i korrespondens med sin bror Graunts tabell och beräkningen av den förväntade livslängden².

Några år innan Neumann kom i kontakt med *Royal Society* beskrev även Leibniz en beräkning av förväntad livslängd i en uppsats som förblev opublicerad. År 1709 var det Nikolaus I Bernoullis tur. År 1725 publicerade Abraham de Moivre en *Avhandling om livräntor*. Han noterade särskilt att priset R_k enkelt kunde beräknas för högre åldrar eftersom formeln (2.2) bara innehöll några få termer. Man kunde sedan använda det bakåtkursiva sambandet

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

vilket är lätt att bevisa utgående från (2.2). Med hjälp av det värde som Halley anger för priset vid 70 års ålder kan man alltså kontrollera³ de andra värdena i tabell 2.2.

²Den förväntade livslängden vid ålder k ges av (2.2) med $i = 0$.

³Det finns sannolikt några fel i tabellen, särskilt när det gäller åldrarna 5 och 15 år.

Efter detta demografieriktade uppehåll återvände Halley till sina huvudsakliga forskningsintressen. Mellan 1698 och 1700 seglade han runt Atlanten för att kartlägga jordens magnetfält. År 1704 blev han professor vid Oxfords universitet. Följande år publicerade han en bok om kometer och förutspådde att kometen från 1682, som Kepler hade observerat 1607, skulle komma tillbaka 1758: Den blev känd som "Halleys komet". Han publicerade också en översättning av Apollonius av Pergas bok om koniska sektioner. År 1720 efterträdde han Flamsteed som kunglig astronom. Han försökte lösa problemet med att bestämma geografiska längdgrader genom observationer av månen, ett problem med stor betydelse vid navigation. Han dog i Greenwich 1742 vid 86 års ålder.

Ytterligare läsning

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, Hoboken, New Jersey (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)

Kapitel 3

Euler och den geometriska befolkningstillväxten (1748–1761)

Euler skrev vid flera tillfällen om befolkningsdynamik. I hans avhandling från 1748, *Introductio in analysin infinitorum*, innehöll kapitlet om exponentialfunktionen fyra exempel på exponentiell tillväxt av en befolkning. År 1760 publicerade han en artikel där han kombinerade denna exponentiella tillväxt med en åldersstruktur för befolkningen. Detta arbete är en föregångare till teorin om ”stabila” befolkningar, som utvecklades under 1900-talet och som spelar en viktig roll inom demografin. År 1761 hjälpte Euler också Süßmilch med den andra upplagan av hans avhandling om demografi. Han utarbetade en intressant modell, som är ett slags variant av Fibonaccis sekvens, men publicerade inte sin detaljerade analys.

Leonhard Euler föddes 1707 i Basel, Schweiz. Hans far var en protestantisk präst. År 1720 började Euler studera vid universitetet. Han fick också privatundervisning i matematik av Johann Bernoulli, en av de mest kända matematikerna i generationen efter Leibniz och Newton. Han blev vän med två av Johann Bernoullis söner: Nikolaus II och Daniel. År 1727 anslöt sig Euler till Daniel vid den nybildade vetenskapsakademien i Sankt Petersburg. Förutom matematik var han också intresserad av fysik och många andra vetenskapliga och tekniska ämnen. År 1741 bjöd kung Fredrik II av Preussen in honom att bli chef för den matematiska sektionen vid vetenskapsakademien i Berlin. Euler publicerade ett betydande antal artiklar och böcker om alla aspekter av mekanik (astronomi, elasticitet, vätskor, fasta ämnen) och matematik (talteori, algebra, oändliga serier, elementära funktioner, komplexa tal, differential- och integralkalkyl, differential- och partiella differentialekvationer, optimering, geometri), men också om demografi. Han var den mest produktiva matematikern under sin tid.

År 1748 publicerade Euler en avhandling på latin med titeln *Introductio in analysin infinitorum*. Han tog upp sex exempel i kapitlet om exponentialer och logaritmer: ett om den matematiska teorin om musikaliska skalor, ett annat om återbetalning av ett lån med ränta och fyra om befolkningsdynamik. I det senare antog Euler att befolkningen P_n år n uppfyller följande villkor



Figur 3.1:
Euler (1707–1783)

$P_{n+1} = (1+x)P_n$ för alla heltal n . Tillväxten x är ett positivt reellt tal. Med utgångspunkt i ett initialtillstånd P_0 ges befolkningen år n av följande formel:

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

Detta kallas geometrisk eller exponentiell tillväxt. I det första exemplet ställs frågan:

”Om befolkningen i en viss region ökar årligen med en trettiondel och det vid en tidpunkt fanns 100 000 invånare, hur ser befolkningen ut efter 100 år?”

Svaret är $P_{100} = (1 + 1/30)^{100} \times 100\,000 \approx 2\,654\,874$. För detta exempel inspirerades Euler av den folkräkning i Berlin som ägde rum 1747 och som gav en uppskattning av befolkningen på 107 224 personer. Hans beräkning visar att en befolkning kan öka mer än tiofaldigt inom ett århundrade. Detta är precis vad som då hade observerats för staden London.

Man bör notera att det är mycket enkelt att beräkna $(1 + 1/30)^{100}$ med en modern fickräknare. Men på Eulers tid var man tvungen att använda logaritmer för att undvika många multiplikationer för hand och få resultatet snabbt. Man beräknar först den decimala logaritmen (i bas 10) av P_{100} . Logaritmens grundläggande egenskap $\log(ab) = \log a + \log b$ visar att

$$\log P_{100} = 100 \log(31/30) + \log(100\,000) = 100(\log 31 - \log 30) + 5.$$

Logaritmen introducerades 1614 av skotten John Napier. Hans vän Henry Briggs publicerade den första tabellen över decimallogaritmer 1617. År 1628 kompletterade holländaren Adriaan Vlacq Briggs arbete genom att

publicera en tabell med decimallogaritmer för heltal från 1 till 100 000 med tiosiffrig precision. Det är denna typ av tabell som Euler använde för att få fram $\log 30 \approx 1,477121255$, $\log 31 \approx 1,491361694$ och slutligen $\log P_{100} \approx 6,4240439$. Det återstår att hitta talet P_{100} vars logaritm är känd. Eftersom decimallogaritmerna för heltal från 1 till 100 000 ligger mellan 0 och 5, söker man i stället logaritmen för $P_{100}/100$, som är 4,4240439. Man kan kontrollera i tabellen över logaritmer att $\log 26548 \approx 4,424031809$ och $\log 26549 \approx 4,424048168$. Genom att ersätta den logaritmiska funktionen med en rät linje mellan 26 548 och 26 549 fick Euler fram att

$$\frac{P_{100}}{100} \approx 26548 + \frac{4,4240439 - 4,424031809}{4,424048168 - 4,424031809} \approx 26548,74.$$

Så $P_{100} \approx 2\,654\,874$.

Det andra exemplet om befolkningsdynamik i Eulers bok är följande:

”Eftersom alla människor efter syndafloden härstammade från en befolkning på sex personer, och om vi antar att befolkningen efter tvåhundra år var 1 000 000, vad var den årliga tillväxten?”

Eftersom $10^6 = (1+x)^{200} \times 6$, får vi med en fickräknare $x = (10^6/6)^{1/200} - 1 \approx 0,061963$. Med logaritmtabeller måste man gå igenom $\log(10^6) = 200 \log(1+x) + \log 6$ för att få $\log(1+x) = (6 - \log 6)/200 \approx 0,0261092$ och $1+x \approx 1,061963$. Euler kunde alltså dra slutsatsen att befolkningen skulle öka med $x \approx 1/16$ per år. För att förstå ursprunget till detta exempel måste man komma ihåg att samtida filosofer började förneka sanningen i de bibliska berättelserna. En bokstavlig tolkning skulle fastställa tidpunkten för syndafloden till cirka 2350 f.Kr. med följande överlevande: Noa, hans tre söner och deras fruar. I Första Moseboken står det:

”Dessa tre var Noas söner, och från dem härstammar alla människor på hela jorden.”

En befolkningstillväxt på 1/16 (eller 6,25 %) per år efter översvämningen verkade inte alltför realistisk för Euler. Eftersom han var son till en protestantisk präst och hade varit religiös hela sitt liv, drog han slutsatsen:

”Därför är det helt löjligt för de otroende att invända att hela jorden inte skulle kunna befolkas på så kort tid med en enda människa i början.”¹.

¹I den bok som Graunt publicerade 1662 (se kapitel 2) finns en liknande anmärkning:

Euler noterade också att om tillväxten hade fortsatt i samma takt fram till 400 år efter floden skulle befolkningen ha varit $(1+x)^{400} \times 6 = (10^6/6)^2 \times 6 \approx 166$ miljarder:

”Hela jorden skulle dock aldrig kunna bära en sådan befolkning.”

Denna idé skulle utvecklas kraftigt av Malthus ett halvt sekel senare (se kapitel 5).

Eulers tredje exempel frågar:

”Om den mänskliga befolkningen fördubblas varje århundrade, vilken är då den årliga tillväxten?”

Eftersom $(1+x)^{100} = 2$ får vi med en fickräknare $x = 2^{1/100} - 1 \approx 0,00695$. Med logaritmtabeller, $100 \log(1+x) = \log 2$. Så $\log(1+x) \approx 0,0030103$ och $1+x \approx 1,00695$. Befolkningen ökar alltså med $x \approx 1/144$ varje år. Det fjärde och sista exemplet ställer samma fråga:

”Om människans befolkning ökar årligen med $1/100$ vill vi veta hur lång tid det tar för befolkningen att bli tio gånger så stor.”

Med $(1+1/100)^n = 10$ finner vi $n \log(101/100) = 1$. Så $n = 1/(\log 101 - 2) \approx 231$ år. Detta är allt man kan hitta i *Introductio in analysin infinitorum* från 1748 om befolkningsdynamik. Euler skulle återkomma till detta ämne mer grundligt några år senare.

År 1760 publicerade han i vetenskapsakademins proceedings i Berlin ett arbete med titeln *En allmän undersökning av dödligheten och förökningen av människosläktet*. Detta arbete var ett slags syntes mellan hans tidigare analys av befolkningarnas geometriska tillväxt och tidigare studier av livstabeller (se kapitel 2). Euler tog till exempel hänsyn till problemet:

”Genom att känna till antalet födselar och begravingar som sker under ett år kan man ta reda på antalet levande och deras årliga ökning, för en given hypotes om dödlighet.”

Euler antog här att följande tal är kända:

- antalet födda B_n under år n ;
- antalet dödsfall D_n under år n ;

”Ett par, nämligen Adam och Eva, som fördubblar sig vart 64:e år under de 5 160 år som är världens ålder enligt Skrifterna, kommer att producera långt fler människor än vad som nu finns i den. Därför är världen inte mer än 100 000 år gammal, som vissa fåfängt inbillar sig, och inte heller mer än vad Skriften säger om den.”

- andelen q_k av de nyfödda som uppnår ålder $k \geq 1$.

Låt P_n vara befolkningen år n . Euler gjorde ytterligare två implicita antaganden:

- Befolkningen ökar geometriskt: $P_{n+1} = rP_n$ (vi fastställer $r = 1 + x$);
- Förhållandet mellan antalet födda och befolkningen är konstant:
 $B_n/P_n = m$.

Dessa två antaganden innebär att antalet födselar ökar geometriskt och i samma takt: $B_{n+1} = rB_n$. Euler betraktade sedan befolkningens tillstånd med hundra års mellanrum, till exempel mellan åren $n = 0$ och $n = 100$, och antog att ingen överlever längre än hundra år. För att förtydliga framställningen kallar vi $P_{k,n}$ ($k \geq 1$) den befolkning som lever i början av år n och som föddes år $n - k$. Kalla $P_{0,n} = B_n$ för antalet födda under år n . Från definitionen av överlevnadskoefficienten q_k har vi $P_{k,n} = q_k P_{0,n-k} = q_k B_{n-k}$. Så

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

Genom att dividera denna ekvation med $r^{100} P_0$ erhåller vi

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

Detta är den ekvation som ibland kallas ”Eulers ekvation” inom demografien. Om vi räknar födselar och dödsfall separat får vi följande:

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

Antalet dödsfall ökar alltså geometriskt: $D_{n+1} = rD_n$. Dessutom,

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

Genom att sätta in detta i ekvation (3.1) kommer vi slutligen fram till ekvationen

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

där det bara finns en okänd kvar: r . Detta är vad som brukar kallas en implicit ekvation eftersom vi inte kan få fram r som en funktion av de andra

parametrarna. Men vi kan beräkna vänster och höger sida av ekvation (3.4) för ett fast värde på r och låta r variera tills de två sidorna är lika stora. Det värde på r som på så sätt erhålls ger befolkningens tillväxttakt $x = r - 1$. Observera att vi från ekvationerna (3.1) och (3.3) får följande uttryck för befolkningen P_n :

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

När befolkningen är stationär ($r = 1$) är detta uttryck detsamma som det som Halley använde för att uppskatta befolkningen i staden Breslau (se kapitel 2).

Euler tog också ställning till följande fråga:

”Hypoteserna om dödlighet och fruktsamhet är givna, om man känner till antalet levande, för att hitta hur många det kommer att finnas vid varje ålder.”

Eftersom överlevnadskoefficienterna q_k och fertilitetskoefficienten m är kända kan tillväxten r beräknas från ekvation (3.1). Under år n är antalet personer som föds år $n - k$ lika med $q_k B_{n-k} = q_k B_n / r^k$ (med $q_0 = 1$). Andelen av den totala befolkningen som är i åldern k är alltså

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

Denna andel är konstant. Med Lotkas terminologi (se kapitel 10) sägs befolkningen vara ”stabil”: Ålderspyramiden behåller samma form genom tiderna.

Euler undersökte sedan på nytt problemet med att konstruera en livslängdstabell när befolkningen inte är stationär utan ökar geometriskt:

”När man känner till antalet levande människor, jämför man antalet födda med antalet döda i varje ålder under ett år för att få fram dödlighetslagen.”

Med ”dödlighetslag” menade Euler överlevnadskoefficienterna q_k . Den totala befolkningen antas nu vara känd genom en folkräkning, vilket inte var fallet för Halley (se kapitel 2). Ekvation (3.2) visar att tillväxten är

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

Låt $D_{k,n}$ vara antalet personer som dör vid en ålder av k under år n : Dessa personer föddes år $n - k$. Så $D_{k,n} = (q_k - q_{k+1}) B_{n-k}$. Men $B_{n-k} =$

B_n/r^k . Överlevnadskoefficienterna q_k kan därför beräknas med hjälp av rekursionsformeln

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

för alla $k \geq 0$, med $q_0 = 1$. Denna formel multiplicerad med B_n ger tillbaka formeln (2.1) som Halley använde för det stationära fallet $r = 1$. Euler insisterade ändå på att hans metod för att beräkna överlevnadskoefficienterna q_k förutsätter att befolkningen ökar regelbundet, med undantag för olyckor som pestepidemier, krig, svält osv. Om folkräkningarna på Eulers tid hade registrerat befolkningens ålder (som i Sverige) skulle detta antagande ha varit onödigt och koefficienterna q_k skulle ha kunnat beräknas lättare.

Med tanke på överlevnadskoefficienterna q_k visade Euler också hur man kan beräkna priset på livräntor. Han nämnde inte Halleys eller de Moivres arbeten i detta ämne. Euler använde en ränta på 5% och den livstabell som publicerades 1742 av holländaren Willem Kersseboom.

Euler var inte den enda vetenskapsmannen som intresserade sig för demografi vid Berlinakademien. Hans kollega Johann Peter Süßmilch hade 1741 publicerat en avhandling på tyska med titeln *Den gudomliga ordningen hos förändringen av det mänskliga släktet: Genom födsel, död och fortplantning av densamma*, som idag anses vara den första avhandlingen som helt och hållet ägnas åt demografi. Süßmilch hade också skrivit en bok *Om den snabba tillväxten i staden Berlin* år 1752.



Figur 3.2:
Süßmilch (1707–1767)

År 1761 publicerade Süßmilch en andra upplaga av sin avhandling. I kapitlet *Om ökningstakten och fördubblingstiden för befolkningar* inkluderade han en intressant matematisk modell som Euler hade utarbetat åt honom. Modellen liknade Fibonaccis modell (se kapitel 1) men för en mänsklig befolkning. Euler utgick från ett par (en man och en kvinna) som

båda var 20 år gamla år 0 och antog att människor dör vid 40 års ålder och gifter sig vid 20 års ålder, samtidigt som varje par får sex barn: två barn (en pojke och en flicka) vid 22 års ålder, ytterligare två vid 24 års ålder och de sista två vid 26 års ålder. Genom att räkna åren två och två så att B_i är antalet födselar under året $2i$, kom Euler fram till att

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

för alla $i \geq 1$. De initiala villkoren motsvarar $B_{-12} = 0$, $B_{-11} = 0$, $B_{-10} = 2$ och $B_i = 0$ för $-9 \leq i \leq 0$. Euler kunde således beräkna antalet födselar, vilket framgår av den andra kolumnen i tabell 3.1. Antalet dödsfall D_i under år $2i$ är då lika med antalet födda under år $2i - 40$: $D_i = B_{i-20}$ för $i \geq 10$ medan $D_i = 0$ för $i \leq 9$. När det gäller antalet P_i av människor som lever år $2i$ är det lika med antalet människor som lever år $2i - 2$, plus antalet födda år $2i$, minus antalet döda år $2i$: $P_i = P_{i-1} + B_i - D_i$.

Detta kapitel i Süßmilchs bok avslutas med en anmärkning som redan kunde ha gjorts om Fibonacci-sekvensen:

”Den stora ordning som tycks råda i Eulers tabell hindrar inte antalet födselar från att följa ett slags utveckling som man kallar rekursiva talföljder [...] Oavsett den initiala ordningen i dessa utvecklingar blir de till en geometrisk utveckling om de inte avbryts och ordningen i början bleknar undan för undan och försvinner nästan helt.”

I boken sägs inte mer om matematiken i denna befolkningsmodell. Euler drev dock studien mycket längre i ett manuskript med titeln *Om människosläktets tillväxt*, som förblev opublicerat under hans livstid. När han sökte efter en lösning på ekvationen (3.5) av formen $B_i = cr^i$, dvs. av formen av en geometrisk talföljd, fick han efter förenkling en polynomisk ekvation av grad 13:

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

Han letade efter en lösning nära $r = 1$ och noterade, med hjälp av en logaritmtabell för beräkning av r^{13} , att

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

Så ekvation (3.6) har en rot mellan 1,09 och 1,10. Genom att uppskatta funktionen $1 + r + r^2 - r^{13}$ med en linje på detta intervall fick Euler följande resultat:

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

Tabell 3.1: Eulers tabell.

i	födda	döda	levande
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	födda	döda	levande
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	1076
62	282	2	1356
63	252	0	1608
64	180	0	1788
65	100	2	1886
66	42	10	1918
67	14	30	1902
68	16	60	1858
69	56	90	1824
70	154	102	1876
71	322	90	2108
72	532	60	2580
73	714	30	3264
74	786	10	4040
75	714	2	4752
76	532	2	5282
77	322	12	5592
78	156	42	5706
79	72	100	5678

i	födda	döda	levande
80	86	180	5584
81	226	252	5558
82	532	282	5808
83	1008	252	6564
84	1568	180	7952
85	2032	100	9884
86	2214	42	12056
87	2032	14	14074
88	1568	16	15626
89	1010	56	16580
90	550	154	16976
91	314	322	16968
92	384	532	16820
93	844	714	16950
94	1766	786	17930
95	3108	714	20324
96	4608	532	24400
97	5814	322	29892
98	6278	156	36014
99	5814	72	41756
100	4610	86	46280
101	3128	226	49182
102	1874	532	50524
103	1248	1008	50764
104	1542	1568	50738
105	2994	2032	51700
106	5718	2214	55204
107	9482	2032	62654
108	13530	1568	74616
109	16700	1010	90306
110	17906	550	107662
111	16702	314	124050
112	13552	384	137218
113	9612	844	145986
114	6250	1766	150470
115	4664	3108	152026
116	5784	4608	153202
117	10254	5814	157642
118	18194	6278	169558
119	28730	5814	192474

Eftersom åren räknas två och två tenderar antalet födselar att multipliceras med \sqrt{r} varje år. Detta antal fördubblas varje n år om $(\sqrt{r})^n = 2$, dvs. varje $n = 2 \log 2 / \log r \approx 15$ år. Eftersom $B_i \approx cr^i$ asymptotiskt sett är lika med B_{i-20} och eftersom antalet D_i dödsfall under år $2i$ är lika med B_{i-20} , får vi $D_i \approx B_i/r^{20}$ med $r^{20} \approx 6,25$. Antalet födda är ungefär sex gånger så stort som antalet döda. Eftersom antalet P_i av människor som lever år $2i$ är lika med $B_i + B_{i-1} + \dots + B_{i-19}$, får vi också följande resultat:

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

Den totala befolkningen är ungefär tio gånger större än antalet födda.

Beviset för att sekvensen (B_i) som visas i tabell 3.1 verkligen växer asymptotiskt som r^i är mer komplicerat. Det var känt sedan Abraham de Moivres arbete om rekursiva talföljder att man genom att införa en "genererande funktion",

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

kan uttrycka $f(x)$ som en rationell funktion. Euler hade förklarat metoden i sin *Introductio in analysin infinitorum* 1748: Den upprepande relationen (3.5) ger faktiskt

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

Så $f(x) = \frac{2x + 2x^2 + 2x^3 + 2x^{12}}{1 - x^{11} - x^{12} - x^{13}}$. Euler visste att en sådan rationell funktion kan delas in i formen

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

talen x_1, \dots, x_{13} är de reella eller komplexa rötterna till ekvationen $1 - x^{11} - x^{12} - x^{13} = 0$. Så

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

Eftersom B_i är koefficienten för x^i i $f(x)$, fick Euler fram att

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

då $i \rightarrow +\infty$, där x_k är roten ur den minsta modulen. Med andra ord tenderar B_i att växa geometriskt som $(1/x_k)^i$. Det återstod att notera att x_k är en rot till ekvationen $1 - x^{11} - x^{12} - x^{13} = 0$ om och endast om $r = 1/x_k$ är en rot till ekvationen (3.6). Vissa detaljer i beviset klargjordes slutligen av Gumbel 1916.

Süssmilch publicerade en tredje upplaga av sin avhandling 1765 och dog i Berlin 1767. Euler, som var på dålig fot med Preussens kung, återvände till Sankt Petersburg 1766. Trots att han förlorade synen fortsatte han att publicera ett stort antal verk med hjälp av sina söner och kollegor, särskilt om algebra, integralkalkyl, optik och skeppsbyggnad. Hans *Bref till en tysk prinsessa, i åtskilliga fysiska och filosofiska ämnen*, som skrevs i Berlin mellan 1760 och 1762, publicerades mellan 1768 och 1772 och blev en bästsäljare i hela Europa. Euler dog i Sankt Petersburg 1783. Hans bidrag till matematisk demografi, särskilt hans analys av den ”stabila” ålderspyramiden i en exponentiellt växande befolkning, skulle återupptäckas först på 1900-talet (se kapitlen 10 och 21).

Ytterligare läsning

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). digzeitschriften.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)

Kapitel 4

Daniel Bernoulli, d'Alembert och smittkoppsympningen (1760)

År 1760 skrev Daniel Bernoulli en artikel om smittkoppor. På hans tid var ympning en omdebatterad metod. Den gav ett visst skydd samtidigt som risken för tidig död var ständigt närvarande. Han använde Halleys tabeller och det som då var känt om smittkoppor för att visa att ympning var fördelaktigt om den förknippade dödsrisken kunde hållas under 11 %. Han argumenterade för att ympning kunde öka den förväntade livslängden vid födseln med upp till tre år. Den första epidemiologiska matematiska modellen presenterades i Bernoullis arbete, som d'Alembert förhöll sig kritisk till.

Daniel Bernoulli föddes år 1700 i Groningen i Nederländerna. Två kända matematiker fanns i familjen: fadern Johann Bernoulli och farbrodern Jakob Bernoulli. År 1705 flyttade Johann till Basel i Schweiz där han tillträdde den professur som ledigförklarades efter Jakobs död. Johann ville inte att hans son skulle studera matematik. Daniel valde i stället medicin och doktorerade 1721 med en avhandling om andning. Han började intressera sig för matematik när han flyttade till Venedig och publicerade en bok 1724 och vann ett pris samma år för en uppsats vid vetenskapsakademien i Paris. Uppsatsen behandlade trimning av timglas vid navigation och han utnämndes efter detta till professor vid Sankt Petersburgs nya akademi. Han arbetade under dessa år med rekursiva talföljder och med Sankt Petersburg-paradoxen. Han återvände 1733 till universitetet i Basel där han efter hand kom att undervisa i botanik, fysiologi och fysik. År 1738 offentliggjordes han den bok i strömmningslära som sedermera blev berömd inom fysiken. Omkring 1753 började han, Euler och d'Alembert samtidigt intressera sig för vibrerande strängar vilket ledde till en viktig matematisk strid.

År 1760 överlämnade han till vetenskapsakademien i Paris ett arbete med titeln *Ett nytt försök till analys av smittkoppornas dödlighet och fördelarna med ympning*. Frågan var om ympning (frivilligt införande av en liten mängd levande smittkoppsvirus i kroppen för att skydda den mot senare infektioner) borde uppmuntras trots att den kunde vara ett dödligt ingrepp. Ympning hade sedan en lång tid tillbaka användts i Asien, och Lady Montagu, hustru till den



Figur 4.1: Daniel Bernoulli
(1700–1782)

brittiske ambassadören i Osmanska riket, införde metoden 1718 i England. I Frankrike övervägde man motvilligt ympning trots att Ludvig XIV:s äldsta son avled i smittkoppor 1711. Voltaire, överlevde smittkoppor 1723 och som levde flera år i exil i England för att följa den senaste utvecklingen, pläderade för ympning i sina *Filosofiska brev* 1734. Även en annan överlevare, den franske vetenskapsmannen La Condamine, rekommenderade ympning vid vetenskapsakademien i Paris 1754.

Innan Maupertuis dog i Basel 1759 uppmuntrade han Daniel Bernoulli att studera ympningsproblemet ur ett matematiskt perspektiv. Utmaningen var att kunna jämföra ympningens långsiktiga fördelar med den omedelbara dödsrisken. Bernoulli förenklade först antagandena enligt följande:

- Personer som smittas med smittkoppor för första gången dör med sannolikheten p (oberoende av ålder) och de överlever med sannolikheten $1 - p$;
- Alla har en sannolikhet q att bli smittade varje år. Sannolikheten att bli smittad för en individ i åldersspannet x till $x + dx$ är $q dx$, där dx är en infinitesimal tidsperiod;
- De som överlever en smittkoppsinfektion får livslång immunitet.

Låt $m(x)$ vara den naturliga dödligheten vid ålder x : Sannolikheten för att en individ dör under ett infinitesimalt tidsintervall dx mellan åldrarna x och $x + dx$ är $m(x) dx$. Betrakta en grupp av P_0 personer som ingår i samma årskull och inför beteckningarna

- $S(x)$ för antalet ”mottagliga” personer som fortfarande lever vid en

ålder av x utan att ha utsatts för smittkoppor¹;

- $R(x)$ för antalet personer som lever vid ålder x och som har överlevt från smittkoppor;
- $P(x) = S(x) + R(x)$ för det totala antalet människor som lever vid ålder x .

Individer föds vid åldern $x = 0$. Så $S(0) = P(0) = P_0$ och $R(0) = 0$. Daniel Bernoulli använde sig av metoder som hade utvecklats i slutet av 1600-talet av Newton, Leibniz och senare av sin far och konstaterade att varje mottaglig individ mellan åldern x och åldern $x + dx$ (med dx oändligt liten) har en sannolikhet $q dx$ att smittas av smittkoppor och en sannolikhet $m(x) dx$ att dö av andra orsaker. Variationen av antalet mottagliga personer är alltså $dS = -Sq dx - Sm(x) dx$. Detta leder till differentialekvationen

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

I denna ekvation betecknar dS/dx derivatan av funktionen $S(x)$. Under samma lilla tidsintervall är antalet människor som dör av smittkoppor $pSq dx$ och antalet människor som överlever av smittkoppor $(1 - p)Sq dx$. Dessutom finns det också $Rm(x) dx$ människor som dör av andra orsaker än smittkoppor. Detta leder till ytterligare en differentialekvation:

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

Addition av de två ekvationerna ger nu

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

Bernoulli kunde med hjälp av (4.1) och (4.3) visa att den kvarvarande andelen av människor som fortfarande är mottagliga vid ålder x ges av

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

¹Detta är de facto ett väntevärde som kan ta alla reella värden inte bara heltalsvärden

Bernoulli härledde sambandet (4.4) genom att eliminera $m(x)$ från ekvationerna (4.1) och (4.3):

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

En ommöblering av uttrycken ger nu att

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

Vi noterar att vänsterledet är derivatan av $f(x) = S(x)/P(x)$ (proportionen av mottagliga personer i befolkningen med åldern x). Därmed fås

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

Lösningen till denna typ av ekvation hade varit känd sedan decennier tillbaka genom Jakob Bernoulli, Daniels farbror. Genom att dividera ekvationen med f^2 och genom att sätta $g(x) = 1/f(x)$ ser vi att $dg/dx = qg - pq$ och att $g(0) = 1/f(0) = 1$. Genom att sätta $h(x) = g(x) - p$ så får vi $dh/dx = qh$. Så $h(x) = h(0)e^{qx} = (1-p)e^{qx}$. Slutligen $g(x) = (1-p)e^{qx} + p$ och $f(x) = 1/g(x)$.

För att tillämpa sin teori så använde Bernoulli Halleys tabell (se kapitel 2). Tabellen anger antalet personer som fortfarande lever i början av år x (med $x = 1, 2, \dots$) i en årskull på 1 238 som föddes under år 0. Men Bernoulli behövde antalet personer $P(x)$ som faktiskt uppnår ålder x och det är inte samma sak. Eftersom Bernoulli – liksom de flesta av sina samtida – inte insåg skillnaden (Halleys artikel är faktiskt inte särskilt tydlig) behöll han siffrorna i Halleys tabell men korrigerade det första talet 1 238, med 1 300 för att erhålla en realistisk dödlighet under det första levnadsåret. Dessa siffror visas i den andra kolumnen i tabell 4.1.

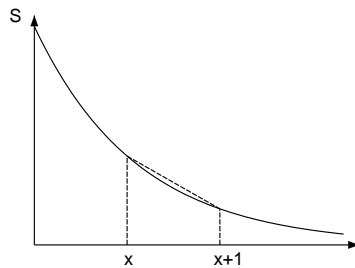
Bernoulli fixerade sannolikheten att dö av smittkoppor till $p = 1/8 = 12,5\%$. Detta stämmer väl överens med samtida observationer. Den årliga sannolikheten för att drabbas av smittkoppor q kunde inte uppskattas direkt. Bernoulli testade därför sannolikt sin modell med flera värden på q innan han slutligen valde det värde som gör att antalet dödsfall på grund av smittkoppor enligt beräkningarna nedan uppgår till ungefär $1/13$ av det totala antalet dödsfall, den andel som då hade observerats i flera europeiska städer. Valet

Tabell 4.1: Halleys tabell och Bernoullis beräkningar.

Ålder x	Levande $P(x)$	Mottaglig $S(x)$	Immun $R(x)$	Smittkopps- död	Inga smittkoppor $P^*(x)$
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
⋮	⋮	⋮	⋮	⋮	⋮

$q = 1/8$ per år visade sig fungera bra, vilket vi också ska se².

Genom sambandet (4.4) och värdena för $P(x)$ i tabellens andra kolonn beräknar vi antalet $S(x)$ av mottagliga personer i åldern x : Detta är tabellens tredje kolonn avrundat till närmaste heltal. Den fjärde kolonnen visar antalet personer $R(x) = P(x) - S(x)$ i åldern x som har överlevt smittkoppor. Den femte kolonnen visar i raden som motsvarar ålder x antalet dödsfall på grund av smittkoppor mellan ålder x och ålder $x + 1$. I teorin ges antalet av integralen $pq \int_x^{x+1} S(t) dt$. Trapetsformeln $pq [S(x) + S(x + 1)]/2$ ger emellertid en bra approximation och vi åskådliggör den i figur 4.2: Trapetsets area ligger nära arean under kurvan som är funktionens integral.



Figur 4.2: Den streckade trapetsens area är en bra approximation för integralen av funktionen S mellan x och $x + 1$.

Bernoulli lade märke till att summan av alla siffror i den femte kolumnen ger 98 dödsfall i smittkoppor före 24 års ålder. Om vi fortsatte tabellen för äldre åldrar skulle vi bara hitta tre ytterligare dödsfall i smittkoppor bland de 32 personer som fortfarande är mottagliga vid 24 års ålder. Totalt 101 personer i en årskull på 1 300 födda dog i smittkoppor. Detta motsvarar nästan precis den förväntade andelen på $1/13$.

Bernoulli föreställde sig sedan en situation där smittkoppor kunde ympas till alla vid födseln på ett sätt som inte orsakade några dödsfall. Frågan var om man kunde uppskatta ökningen av den förväntade livslängden efter en utrotning av sjukdomen. Låt oss utgå från samma antal födselar P_0 och låt oss kalla $P^*(x)$ antalet människor i åldern x i en omgivning där smittkopporna är utrotade. Då gäller

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

Bernoulli visade även att

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

²Att p och q är lika stora är bara en tillfällighet.

där $P(x)$ är befolkningen i åldern x enligt ovan i en omgivning där smittkoppor finns.

Precis som ovan kan $m(x)$ elimineras i ekvationerna (4.6) och (4.3). Bernoulli erhöll efter en ommöblering av uttrycken

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

Han införde $h(x) = P(x)/P^*(x)$. Med hjälp av sambandet (4.4) multiplicerade han täljaren och nämnaren med e^{-qx} och fick

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

Detta är likvärdigt med $\frac{d}{dx} \log h = \frac{d}{dx} \log(1 - p + pe^{-qx})$, där log här står för den naturliga logaritmen och inte för tiologaritmen. Men $h(0) = 1$. Därmed är $h(x) = 1 - p + pe^{-qx}$.

Observera att förhållandet $P(x)/P^*(x)$ närmar sig $1 - p$ när åldern x blir tillräckligt hög. Den sjätte kolonnen i tabell 4.1 visar $P^*(x)$. Ett sätt att jämföra $P(x)$ och $P^*(x)$ är att uppskatta den förväntade livslängden vid födseln. Med smittkoppor fås då

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

Ett liknande uttryck där $P(x)$ ersätts av $P^*(x)$ gäller i en omgivning utan smittkoppor. Bernoulli använde trapetsformeln

$$\left[\frac{1}{2}P(0) + P(1) + P(2) + \dots \right] / P_0$$

(figur 4.2). Genom att utvidga tabellen från 24 år till 84 år (se tabell 2.1) fick han slutligen fram att en ungefärlig förväntad livslängd E i en omgivning med smittkoppor antar värdet $[\frac{1}{2} \times 1300 + 1000 + \dots + 20] / 1300 \approx 26,57$ år, det vill säga 26 år och 7 månader. Om smittkopporna kunde utrotas så uppgick motsvarande ungefärliga förväntade livslängd E^* till $[\frac{1}{2} \times 1300 + 1015 + \dots + 23] / 1300 \approx 29,65$ år, dvs. 29 år och 8 månader. Ympning vid födseln skulle alltså öka den förväntade livslängden med mer än tre år.

Vi konstaterar att det finns enklare och snabbare metoder än de som Bernoulli använde för att förstå dessa samband. Om vi utgår från differentialekvationen (4.1) för $S(x)$ så ser vi först att

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

Substitution av detta uttryck i ekvation (4.2) för $R(x)$ ger att

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

Ekvation (4.6) för $P^*(x)$ visar att

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

Sambandena (4.4) och (4.7) följer därur!

Det är naturligtvis inte riskfritt att ympa med en mindre sjukdomsalstrande stam av smittkoppor. Om p' är sannolikheten för att dö av smittkoppor strax efter ympning ($p' < p$), så blir den förväntade livslängden $(1 - p')E^*$ när alla ympas vid födseln. Den förväntade livslängden är högre än den naturliga livslängden E om $p' < 1 - E/E^*$ eller cirka 11 %. Uppgifter om p' fanns inte tillgängliga vid den tiden. Men Bernoulli uppskattade dock risken p' till mindre än 1 %. För honom var det ingen tvekan: Ympning måste främjas av staten. Han drog slutsatsen:

”Jag önskar helt enkelt att man i en fråga som är så viktig för välfärden inte fattar några beslut i frånvaro av all den kunskap som lite analys och beräkningar kan tillhandahålla.”

Bernoullis arbete presenterades vid vetenskapsakademien i Paris i april 1760. I november lade d'Alembert fram en kommentar med titeln *Om tillämpningen av sannolikhetsteorin vid ympning mot smittkoppor*. Kommentaren publicerades kort därefter i den andra volymen av hans *Opuscles mathématiques* med mer detaljerade beräkningar och tillsammans med ett annat arbete med titeln *Matematisk teori om ympning*. D'Alembert kritiserade Bernoullis antaganden om att sannolikheten för infektion och sannolikheten för att dö av smittkoppor skulle vara oberoende av ålder. Han föreslog i stället en lösning som inte kräver sådana antaganden. Kalla $v(x)$ dödligheten på grund av smittkoppor vid ålder x , $m(x)$ dödligheten på grund

av andra orsaker och $P(x)$ antalet människor som fortfarande lever. Då gäller

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

Genom att jämföra med ekvation (4.3) ser vi att $v(x) = pqS(x)/P(x)$. Här får vi

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

där $P^*(x)$ står för antalet människor som lever vid ålder x när smittkopporna har försvunnit.



Figur 4.3:
D'Alembert (1717–1783)

Vi kan antingen ersätta funktionen $m(x)$ mellan ekvationerna (4.6) och (4.9) eller använda formeln (4.8) för $P^*(x)$ och notera att lösningen av (4.9) ges av

$$P(x) = P_0 \exp\left(-\int_0^x [v(y) + m(y)] dy\right).$$

Formeln (4.10) som gavs av d'Alembert innehåller ingen motsägelse till Bernoullis formel (4.7). Men den använder informationen $v(x)$, som inte var tillgänglig vid den tidpunkten eftersom dödsregistren bara innehöll dödsorsaken och inte offrets ålder. D'Alembert menade att man inte kunde dra några slutsatser om ympningens effekter innan sådana uppgifter fanns tillgängliga.

D'Alembert ifrågasatte även användbarheten av den förväntade livslängden som kriterium. Alla år viktades ju nu lika oberoende av om de ligger nära eller långt fram i tiden. Han konstaterade att alla åldrar vare sig

ur individens eller statens synvinkel kan vara ”likvärdiga”. Unga och gamla individer måste ha mindre värde än individer i medelåldern. D’Alembert tog trots sin kritik emellertid ställning för ympning.

Utgivningen av Bernoullis arbete försenades och offentliggjordes först 1766. D’Alembert lyckades emellertid få sitt eget arbete publicerat mycket snabbt. Bernoulli uttryckte bitterhet i ett brev till Euler:

”Vad säger du om den store d’Alemberts enorma banaliteter om sannolikheter: Eftersom jag alltför ofta känner mig själv orättvist behandlad i hans publikationer har jag redan för en tid sedan bestämt mig för att inte längre läsa något som kommer ur hans penna. Jag har fattat detta beslut med anledning av ett manuskript om ympning som jag skickade till akademien i Paris för åtta år sedan. Analysen i detta manuskript uppskattades mycket på grund av sitt nyhetsvärde. Jag vågar hävda, att det var som att tillföra ett nytt organ i matematikens kropp. Det verkar som om framgången med min analys orsakade honom hjärtesorg. Han har kritiserat den på tusen olika sätt som alla är lika löjliga och efter att ha fått den väl kritiserad låtsas han vara den första författaren till en teori som han inte bara hört nämnas. Han visste emellertid att mitt manuskript kunde komma ut först efter cirka sju eller åtta år. Han kunde bara ha kännedom om det i sin egenskap av ledamot av akademien. Mitt manuskript borde i detta avseende ha förblivit heligt tills det offentliggjordes. *Dolus an virtus quis in hoste requirat?*”³

Trots Bernoullis och d’Alemberts arbeten inleddes inte någon storskalig ympning i Frankrike. Kung Ludvig XV dog av smittkoppor 1774. Kort därefter ympade hovets läkare resten av kungafamiljen. Problemet förlorade sin betydelse när Edward Jenner upptäckte att ympning av kokoppor på människor (”vaccination”) var ett säkert sätt att erhålla skydd mot smittkoppor. Hans arbete, *En undersökning av orsakerna till och effekterna av variolae vaccina.*, publicerades 1798. Vaccinationsmetoden spred sig nu snabbt över hela Europa. De metoder som utvecklades för att beräkna ökningen av den förväntade livslängden när en dödsorsak avlägsnas används fortfarande i dag.

Under de följande årtiondena började man få tillgång till uppgifter om åldern då människor dog av smittkoppor. Detta ledde till att problemet togs ånyo till behandling, bland annat av

³Vad spelar det för roll om vi besegrar fienden genom tapperhet eller genom list? Vergil: *Aeneid*, bok II.

- Johann Heinrich Lambert, en matematiker vid Berlins akademi, år 1772;
- Emmanuel-Étienne Duvallard, då ansvarig för befolkningsstatistik vid inrikesministeriet i Paris, skrev i sin rapport *Analys och tabeller över smittkopornas inverkan på dödligheten vid varje ålder* (1806);
- Pierre-Simon Laplace i sin *Analytisk teori om sannolikhet* (1812).

Duvillard och Laplace visade bland annat hur formeln (4.7) ändras när parametrarna p och q beror på åldern:

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

Här är $p(x)$ sannolikheten att dö av smittkoppor om man är smittad vid åldern x och $q(x)$ sannolikheten att bli smittad av smittkoppor vid åldern x .

Efter detta arbete med smittkoppor engagerade sig Daniel Bernoulli inte i några andra populationsdynamiska problem. Han avled i Basel 1782 och d'Alambert avled i Paris ett år senare.

Ytterligare läsning

1. D'Alembert, J.: Sur l'application du calcul des probabilités à l'inoculation de la petite vérole. In: *Opuscules 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. Duvallard, 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 (översättare: Stolpe, J.): *Filosofiska brev*. Carmina, Uppsala (1986). gallica.bnf.fr

Kapitel 5

Malthus och hindren för den geometriska tillväxten (1798)

År 1798 publicerade Malthus *En avhandling om befolkningslagen*, där han hävdade att tillgången på livsmedel under en längre tidsperiod inte kunde följa den naturliga tendensen hos människans befolkning att växa exponentiellt. Om befolkningen förblev relativt konstant berodde detta på att en stor del av mänskligheten led av livsmedelsbrist. Malthus såg "befolkningsprincipen" som ett argument mot Godwins och Condorcets skrifter, som betonade framsteg i de mänskliga samhällena. Malthus uppsats påverkade Darwins och Wallaces evolutionsteori och kritiserades av Marx, men omsattes i praktiken med den kinesiska ettbarnspolitiken.

Thomas Robert Malthus föddes 1766 i närheten av London som det sjätte av sju barn. Hans far, en vän och beundrare av Jean-Jacques Rousseau, var hans första lärare. År 1784 började den unge Malthus studera matematik vid universitetet i Cambridge. Han fick sin examen 1791, blev docent vid *Jesus College* 1793 och anglikansk präst 1797.



Figur 5.1: Malthus (1766–1834)

År 1798 publicerade Malthus anonymt en bok med titeln *En avhandling om befolkningslagen: och dess inverkan på samhällets framtida förbättring*

med anmärkningar om herr Godwins, herr Condorcets och andra skriftställares spekulationer. Boken kom som en reaktion mot Godwins *Den politiska rättvisan och dess inflytande på moralen och lyckan* (1793) och Condorcets *Människosläktets andliga förkovran* (1794). Trots de favor som franska revolutionen gjorde i framstegets namn hävdade de två författarna att samhällets framsteg var oundvikliga. Malthus delade inte samma optimism. Han hävdade också att de engelska fattiglagarna, som hjälpte fattiga familjer med många barn, gynnade befolkningstillväxten utan att uppmuntra en liknande ökning av livsmedelsproduktionen. Det föreföll honom som om dessa lagar egentligen inte hjälpte de fattiga, utan tvärtom tycktes en del av samhället vara dömt till elände, hunger eller epidemier, eftersom befolkningen generellt tenderar att alltid växa snabbare än livsmedelsproduktionen. Detta är de gissel som bromsar befolkningstillväxten och som enligt Malthus åsikt är de främsta hindren för samhällets utveckling. Alla teorier som lovar framsteg skulle bara vara utopiska. Dessa idéer fick Malthus att publicera sin bok 1798. Så här sammanfattade han sin tes:

[...] ”Befolkningens kraft är oändligt mycket större än jordens kraft att producera människans försörjning. Befolkningen ökar okontrollerat i ett geometriskt förhållande. Livsmedelsförsörjningen ökar endast i ett aritmetiskt förhållande. En liten bekantskap med siffror kommer att visa hur enorm den första kraften är i jämförelse med den andra. Genom den lag i vår natur som gör mat nödvändig för människans liv måste effekterna av dessa två ojämna krafter hållas lika stora. Detta innebär en stark och ständigt verksam kontroll av befolkningen på grund av svårigheten att försörja sig. Denna svårighet måste drabba någonstans, och måste nödvändigtvis kännas hårt av en stor del av mänskligheten.”

Malthus bok var mycket framgångsrik. Den innehöll få numeriska data. Malthus noterade till exempel att befolkningen i USA hade fördubblats vart tjugofemte år under 1700-talet. Han försökte inte riktigt omsätta sina teser i matematiska modeller men banade väg för senare arbeten av Adolphe Quetelet och Pierre-François Verhulst, som kommer att behandlas i nästa kapitel.

Efter publiceringen av sin bok reste Malthus med vänner först till Tyskland, Skandinavien och Ryssland och sedan till Frankrike och Schweiz. Han sammanställde den information som han samlat in under sina resor och publicerade i sitt namn en mycket kraftigt utökad andra upplaga 1803, med

en annan undertitel: *En avhandling om befolkningslagen, eller en bild av dess tidigare och nuvarande effekter på människans lycka, med en undersökning av våra utsikter när det gäller att i framtiden avlägsna eller mildra de onda effekter som den orsakar.* I denna nya upplaga diskuterades i detalj de hinder som finns för befolkningstillväxten i olika länder: försenade äktenskap, abort, barnamord, svält, krig, epidemier, ekonomiska faktorer... För Malthus var fördröjda äktenskap det bästa alternativet för att stabilisera befolkningen. Fyra andra upplagor av boken följde 1806, 1807, 1817 och 1826. År 1805 blev Malthus professor i historia och politisk ekonomi i en ny skola som Västindiska kompaniet inrättade för sina anställda. Han publicerade också *En undersökning av hyresrättens natur och utveckling* (1815) och *Principer för politisk ekonomi* (1820). År 1819 valdes Malthus in i *Royal Society*. År 1834 var han en av grundarna av *Statistical Society*. Han dog nära Bath samma år.

Malthus arbete hade ett starkt inflytande på utvecklingen av evolutionsteorin. Charles Darwin, som var tillbaka från sin resa ombord på *Beagle*, läste Malthus bok om befolkning 1838. Här är vad han skrev i inledningen till sin berömda bok *Om arternas uppkomst genom naturligt urval*, som publicerades 1859:

”I nästa afdelning afhandlas kampen för tillvaron emellan de organiska varelserna, hvilken ovederläggligt framgår ur förökningens fortgång i geometrisk progression. Det är Malthus lära tillämpad på hela växt- och djurriket.”

Alfred Russel Wallace, som utvecklade evolutionsteorin samtidigt som Darwin, sade också att han fick sina idéer efter att ha läst Malthus bok.

I kontrast till detta står Karl Marx syn på Malthus boks framgång, som kan läsas i en fotnot i hans *Kapitalet*:

”Om läsaren skulle komma att tänka på Malthus, vars *En uppsats om befolkningsprincipen* utkom 1798, vill jag påminna om att denna skrift i sin första form inte är något annat än ett omoget, ytligt och i predikoton framfört plagiat från Defoe, Sir James Stewart, Townsend, Franklin, Wallace m. fl. och inte innehåller en enda självständig tanke. Det rabalder denna pamflett åstadkom, berodde uteslutande på partiintressen. Den franska revolutionen hade fått lidelsefulla försvarare i England. ’Befolkningsprincipen’, som bit för bit formulerades under loppet av det 18:e århundradet, för att sedan under en social kris proklamerades med pukor och trumpeter som det ofelbara motgiftet mot Condorcets och andras läror, hälsades med jubel av den engelska oligarkin som det effektivaste vapnet mot alla

tendenser till mänskliga framsteg. Malthus, högeligen förvånad över sin succé, tog sig då för att i den gamla ramen stoppa in en massa gammalt, planlöst sammanrafsat material samt tillfoga en del nytt, som han inte hade utarbetat själv utan 'lånat' från andra.”

Malthus teser var förvisso inte helt nya. Till exempel tillskrivs han ofta idén att befolkningen tenderar att växa geometriskt, trots att vi i kapitel 3 såg att Euler redan var bekant med denna idé ett halvt sekel tidigare¹. Malthus gav den dock publicitet genom att på ett polemiskt sätt koppla den till verkliga lagstiftningsproblem. Ironiskt nog var det i det kommunistiska Kina som Malthus förslag om att begränsa födslarna skulle finna sin mest slående tillämpning (se kapitel 25).

Ytterligare läsning

1. Condorcet (översättare: Klum, A.): *Människosläktets andliga förkovran*. Carlsson, Stockholm (2002). gallica.bnf.fr
2. Darwin, C. (översättare: Selling, A. M.) *Om arternas uppkomst*. Stockholm (1871). sv.wikisource.org
3. Godwin, W. (översättare: Carlquist, E.): *Den politiska rättvisan och dess inflytande på moralen och lyckan*. Umeå (2017). archive.org
4. Malthus, T.R. (översättare: Backelin, E.): *En avhandling om befolkningslagen*. Daidalos, Göteborg (2012). econlib.org
5. Marx, K. (översättare: Bohman, I.): *Kapitalet : kritik av den politiska ekonomin. Bok 1, Kapitalets produktionsprocess*, Lund (1969) 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 (se kapitlen 14 och 20) skulle kalla ”malthusianska parametern” för befolkningens tillväxttakt. Malthus nämnde Süßmilchs avhandling i sin egen bok.

Kapitel 6

Verhulst och den logistiska ekvationen (1838)

År 1838 introducerade den belgiske matematikern Verhulst den logistiska ekvationen, som är en slags generalisering av ekvationen för exponentiell tillväxt, men med ett maximalt värde för befolkningen. Han använde data från flera länder, särskilt Belgien, för att uppskatta de okända parametrarna. Verhulsts arbete återupptäcktes först på 1920-talet.

Pierre-François Verhulst föddes 1804 i Bryssel. Han disputerade i matematik vid universitetet i Gent 1825. Han var också intresserad av politik. När han var i Italien för att bota sin tuberkulos pläderade han utan framgång för en konstitution för påvestaten. Efter revolutionen 1830 och Belgiens självständighet publicerade han en historisk essä om en patriot från 1700-talet. År 1835 blev han professor i matematik vid det nybildade Fria universitetet i Bryssel.



Figur 6.1:
Verhulst (1804–1849)

Samma år 1835 publicerade hans landsman Adolphe Quetelet, statistiker och chef för observatoriet i Bryssel, *En avhandling om människan och utvecklingen av hennes förmågor*. Quetelet menade att befolkningarna inte kunde växa geometriskt under en längre tidsperiod eftersom de hinder som Malthus nämner utgjorde ett slags "motstånd", som han menade (i analogi med mekaniken) var proportionell mot kvadraten på befolkningstillväxtens

hastighet. Denna analogi hade ingen förankring i verkligheten, men den inspirerade Verhulst.

Verhulst publicerade faktiskt 1838 artikeln *Anmärkingar om lagen om befolkningstillväxt*. Här är några utdrag:

”Vi vet att den berömde Malthus visade principen att människans befolkning tenderar att växa i en geometrisk utveckling så att den fördubblas efter en viss tid, till exempel vart tjugofemte år. Detta är ett obestriddigt påstående om man bortser från den ökande svårigheten att hitta mat [...]

Befolkningens virtuella ökning begränsas alltså av landets storlek och fertilitet. Som ett resultat av detta kommer befolkningen allt närmare ett stabilt tillstånd.”

Verhulst insåg förmodligen att Quetelet’s mekaniska analogi inte var rimlig och föreslog istället följande (fortfarande något godtyckliga) differentialekvation för befolkningen $P(t)$ vid tiden t :

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

När befolkningen $P(t)$ är liten jämfört med parametern K får vi den ungefärliga ekvationen

$$\frac{dP}{dt} \approx rP,$$

vars lösning är $P(t) \approx P(0)e^{rt}$, dvs. exponentiell tillväxt¹. Tillväxten minskar när $P(t)$ närmar sig K . Den skulle till och med bli negativ om $P(t)$ kunde överstiga K . För att få det exakta uttrycket för lösningen av ekvation (6.1) kan vi gå tillväga som Daniel Bernoulli för ekvation (4.5).

Genom att dividera ekvation (6.1) med P^2 och sätta $p = 1/P$ får vi $dp/dt = -rp + r/K$. Med $q = p - 1/K$ får vi $dq/dt = -rq$ och $q(t) = q(0)e^{-rt} = (1/P(0) - 1/K)e^{-rt}$. Vi kan alltså härleda $p(t)$ och $P(t)$.

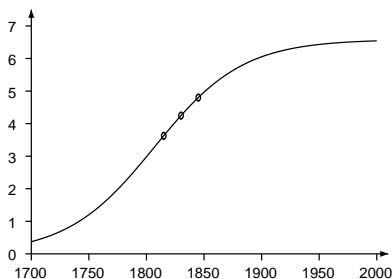
Slutligen får vi efter en enkla beräkningar följande

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

Den totala befolkningen ökar successivt från $P(0)$ vid tiden $t = 0$ till gränsen K , som nås först när $t \rightarrow +\infty$ (figur 6.2).

¹Man brukar tala om geometrisk tillväxt i diskreta tidsmodeller och om exponentiell tillväxt i kontinuerliga tidsmodeller, men det är i princip samma sak.

Figur 6.2: Belgiens befolkning (i miljoner) och den logistiska kurvan. Datapunkterna motsvarar åren 1815, 1830 och 1845. Parametervärdena är de som anges i artikeln från 1845.



Utan att ange vilka värden han använde för de okända parametrarna r och K jämförde Verhulst sitt resultat med uppgifter om befolkningen i Frankrike mellan 1817 och 1831, i Belgien mellan 1815 och 1833, i grevskapet Essex i England mellan 1811 och 1831 och i Ryssland mellan 1796 och 1827. Det visade sig att det stämde ganska bra.

År 1840 blev Verhulst professor vid den kungliga militära skolan i Bryssel. Året därpå publicerade han en *Elementär behandling av elliptiska funktioner* och valdes in i Belgiens kungliga akademi. År 1845 fortsatte han sina befolkningsstudier med en artikel med titeln *Matematiska undersökningar om lagen om befolkningstillväxt*. Han vände sig först tillbaka till Malthus anmärkning enligt vilken befolkningen i USA hade fördubblats vart 25:e år (tabell 6.1).

Tabell 6.1: Officiella folkräkningar i USA.

år	befolkning	år	befolkning
1790	3 929 827	1820	9 638 131
1800	5 305 925	1830	12 866 020
1810	7 239 814	1840	17 062 566

Om vi beräknar förhållandet mellan befolkningen år $n + 10$ och befolkningen år n finner vi 1,350, 1,364, 1,331, 1,335 och 1,326, vilket är ganska konstant. Befolkningen multiplicerades alltså i genomsnitt med 1,34 vart tionde år och med $1,34^{25/10} \approx 2,08$ vart 25:e år. Den hade alltså fortsatt att fördubblas vart 25:e år sedan Malthus uppsats, nästan ett halvt sekel tidigare. Verhulst tillade dock:

”Vi kommer inte att insistera på hypotesen om geometrisk progression, eftersom den bara kan gälla under mycket

speciella omständigheter, till exempel när ett bördigt område av nästan obegränsad storlek råkar vara bebott av människor med en avancerad civilisation, vilket var fallet med de första amerikanska kolonierna.”

I sin artikel återkom Verhulst också till ekvationen (6.1), som han kallade ”logistisk”. Han noterade att kurvan $P(t)$ ökar med en positiv krökning (den är konvex) så länge $P(t) < K/2$ och fortsätter sedan att öka mot K men med en negativ krökning (den är konkav) så snart $P(t) > K/2$. Kurvan har alltså formen av en förvrängd bokstav S (figur 6.2).

Faktum är att $d^2P/dt^2 = r(1 - 2P/K)dP/dt$. Så $d^2P/dt^2 > 0$ om $P < K/2$ och $d^2P/dt^2 < 0$ om $P > K/2$.

Verhulst förklarade också hur parametrarna r och K kan uppskattas utifrån befolkningen $P(t)$ under tre olika, men jämnt fördelade år. Om P_0 är befolkningen vid tiden $t = 0$, P_1 den vid tiden $t = T$ och P_2 den vid tiden $t = 2T$, så visar en omständlig beräkning med utgångspunkt i ekvation (6.2) att

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

Med hjälp av uppskattningarna av Belgiens befolkning för åren 1815, 1830 och 1845 (3,627, 4,247 respektive 4,801 miljoner) fick han fram $K = 6,584$ miljoner och $r = 2,6\%$ per år. Han kunde sedan använda ekvation (6.2) för att förutsäga att Belgiens befolkning skulle vara 4,998 miljoner i början av år 1851 och 6,064 miljoner i början av år 1900 (figur 6.2). Verhulst gjorde en liknande studie för Frankrike. Han fick fram $K = 39,685$ miljoner och $r = 3,2\%$ per år. Eftersom Belgiens och Frankrikes befolkningar under tiden vida överskridit dessa värden på K , ser vi att den logistiska ekvationen kan vara en realistisk modell endast för tidsperioder på några decennier, som i Verhulsts artikel från 1838, men inte för längre tidsperioder.

År 1847 publicerades *Andra undersökningen om lagen om befolkningstillväxt* där Verhulst gav upp den logistiska ekvationen och valde istället en differentialekvation som kan skrivas i formen

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

Han trodde att denna ekvation skulle gälla när befolkningen $P(t)$ är över en viss tröskel. Lösningen är $P(t) = K + (P(0) - K)e^{-rt/K}$. Med hjälp av samma demografiska uppgifter för Belgien uppskattade Verhulst parametrarna r och

K på nytt. Den här gången fann han $K = 9,4$ miljoner för den maximala befolkningen. Det visar hur mycket resultatet kan bero på valet av modell!

Verhulst blev president för Belgiens kungliga akademi 1848, men dog året därpå i Bryssel, troligen av tuberkulos. Trots Verhulsts tveksamhet mellan modellekvationer återinfördes den logistiska ekvationen oberoende flera decennier senare av olika personer. Robertson använde den 1908 för att modellera den individuella tillväxten hos djur, växter, människor och kroppsorgan. McKendrick och Kesava Pai använde den 1911 för tillväxt av populationer av mikroorganismer. Pearl och Reed använde den 1920 för att beräkna befolkningstillväxten i USA, som hade börjat avta. År 1922 uppmärksammade Pearl slutligen Verhulsts arbete. Från och med då inspirerade den logistiska ekvationen många arbeten (se kapitlen 13, 20 och 24). Den maximala populationen K blev så småningom känd som ”bärförmåga” eller ”bärandekapacitet”.

Ytterligare läsning

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 microorganisms: 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 XIX^e 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

Kapitel 7

Bienaymé, Cournot och släktnamnens utdöende (1845–1847)

Den franske statistikern Bienaymé förstod 1845 hur man kan beräkna sannolikheten för att ett släktnamn ska dö ut om varje man har ett antal söner som följer en viss sannolikhetsfördelning. Om det genomsnittliga antalet söner är mindre än eller lika med ett kommer familjenamnet att dö ut. Om genomsnittet är större än ett, är sannolikheten för utdöende strikt mindre än ett. Beviset för hans resultat publicerades två år senare i en bok skriven av hans vän Cournot. Dessa arbeten återupptäcktes först nyligen.

Irenée Jules Bienaymé föddes 1796 i Paris. Han studerade vid *École Polytechnique* och gjorde karriär inom finansministeriet och nådde den höga nivån som generalinspektör. Bienaymé påverkades av Laplaces bok *Analytisk teori om sannolikhet* och fann också tid att publicera artiklar om många tillämpningar av sannolikhetsteorin, t. ex. demografisk och medicinsk statistik (spädbarnsdödlighet, antal födda, förväntad livslängd), sannolikheten för fel i rättsväsendet, försäkringsteori och röstningssystemens representativitet.



Figur 7.1: Bienaymé (1796–1878) och Cournot (1801–1877)

År 1845 skrev Bienaymé en kort notis *Om multiplikationslagen och familjernas varaktighet*, som publicerades i en bulletin från *Société Philomatique* i Paris. Ett antal författare hade redan skrivit om detta ämne. I den andra upplagan av *En avhandling om befolkningslagen* (1803) inkluderade Malthus ett kapitel om befolkningen i Schweiz och konstaterade att

”I staden Bern hade det suveräna rådet från år 1583 till 1654 godkänt 487 familjer i bourgeoisien, varav 379 dog ut under två århundraden och år 1783 återstod endast 108 av dem.”

År 1842 hävdade Thomas Doubleday mer allmänt att överklassfamiljer från adeln eller bourgeoisien hade en större tendens att försvinna än underklassfamiljer. Liknande idéer framfördes i Frankrike av Émile Littré 1844 i en introduktionstext till Auguste Comtes positivistiska filosofi och av Benoiston de Châteauneuf – en vän till Bienaymé – som 1845 publicerade en essä *Om adelsfamiljernas varaktighet i Frankrike*.

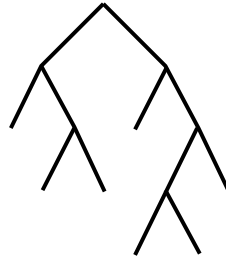
Det var i detta sammanhang som Bienaymé försökte förklara hur det kan komma sig att befolkningen i ett land tenderar att växa geometriskt samtidigt som ett stort antal familjer försvinner. För att angripa detta problem tog han ett förenklat fall där alla män skulle ha samma sannolikhet att ha 0, 1, 2, 3... söner som når vuxen ålder. Närmare bestämt frågade han sig vad sannolikheten var för en man att ha en avkomma som bär hans namn efter n generationer. Om det genomsnittliga antalet söner är mindre än ett, är det uppenbart att denna sannolikhet bör gå mot noll när n växer mot oändligheten. Bienaymé noterade att samma slutsats skulle förbli sann om medelantalet söner var exakt ett¹, t. ex. om det finns en sannolikhet $1/2$ att inte ha någon son och en sannolikhet $1/2$ att ha två söner (figur 7.2). Men i det fallet går sannolikheten att få avkomma i generation n långsammare mot noll: I exemplet skulle den fortfarande vara 5% efter 35 generationer, dvs. efter elva eller tolv århundraden om det går tre generationer per århundrade². Bienaymé noterade slutligen att även om det genomsnittliga antalet söner är större än ett så är det inte säkert att släktlinjen dör ut: Sannolikheten för detta kan beräknas genom att lösa en algebraisk ekvation.

Bienaymés artikel innehöll ingen mer förklaring. År 1847 inkluderade hans vän Antoine-Augustin Cournot, en matematiker och ekonom, några detaljer i en bok med titeln *Om ursprunget och gränserna för korrespondensen mellan algebra och geometri*. Han presenterade problemet i

¹Utom om varje man har exakt en son.

²Som vi kommer att se nedan är denna sannolikhet lika med $1 - x_{35}$ med $x_{n+1} = \frac{1}{2} + \frac{1}{2}x_n^2$ och $x_0 = 0$.

Figur 7.2: Artificiellt exempel på ett släkträd. Förfadern står högst upp i trädet. I varje generation har männen en sannolikhet $1/2$ att inte få någon son och en sannolikhet $1/2$ att få två söner.



form av ett slumpspel men erkände att det var identiskt med Bienaymés studie av släktnamnens utdöende. Om vi behåller tolkningen i termer av släktnamn, betraktade Cournot först det specialfall där män får högst två söner, där p_0 , p_1 och p_2 är sannolikheten för att få 0, 1 respektive 2 söner. Naturligtvis är $p_0 + p_1 + p_2 = 1$. Om man utgår från en förfader är sannolikheten för utdöende efter bara en generation, kalla den x_1 , naturligtvis lika med p_0 . Sannolikheten för utdöende inom två generationer är $x_2 = p_0 + p_1 x_1 + p_2 x_1^2$: Antingen var familjen utdöd redan i första generationen (sannolikhet p_0), eller så fanns det bara en son i första generationen som inte fick någon son (sannolikhet $p_1 x_1$), eller så fanns det två söner i första generationen och ingen av dem fick någon son (sannolikhet $p_2 x_1^2$). Mer allmänt är sannolikheten för utdöende inom n generationer

$$x_n = p_0 + p_1 x_{n-1} + p_2 (x_{n-1})^2.$$

Om det till exempel finns två söner i den första generationen (sannolikhet p_2) kommer familjen att dö ut $n - 1$ generationer senare (dvs. i generation n) med en sannolikhet som är lika med $(x_{n-1})^2$. Cournot noterade att x_n är en ökande sekvens med $x_n \leq 1$ för alla n . Så x_n har en gräns $x_\infty \leq 1$, som är en lösning på ekvationen

$$x = p_0 + p_1 x + p_2 x^2.$$

Med hjälp av $p_1 = 1 - p_0 - p_2$ är denna ekvation likvärdig med $0 = p_2(x - 1)(x - p_0/p_2)$. Det finns alltså två rötter: $x = 1$ och $x = p_0/p_2$. Tre fall kan särskiljas beroende på det genomsnittliga antalet söner $p_1 + 2p_2$, som också är lika med $1 - p_0 + p_2$ och som vi ska kalla \mathcal{R}_0 . Om $\mathcal{R}_0 < 1$ är $p_0/p_2 > 1$, så $x = 1$ är det enda möjliga värdet för gränsen x_∞ . Familjenamnet kommer säkert att dö ut. Om $\mathcal{R}_0 = 1$ är båda rötterna lika med 1 och slutsatsen är densamma. Om $\mathcal{R}_0 > 1$, argumenterade Cournot att x_∞ borde vara lika med den andra roten p_0/p_2 , eftersom sannolikheten för utdöende uppenbarligen måste vara 0 i det speciella fallet där $p_0 = 0$.

Cournot nämnde kort det mer allmänna fallet där män kan få högst m söner med sannolikheterna p_0, p_1, \dots, p_m . Slutsatsen beror på samma sätt på värdet av $\mathcal{R}_0 = p_1 + 2p_2 + \dots + mp_m$, det genomsnittliga antalet söner, i förhållande till 1. Ekvationen för x_∞ , som är $x = p_0 + p_1x + \dots + p_mx^m$, har alltid roten $x = 1$. Den har endast en annan positiv rot, vilken ger sannolikheten för utdöende x_∞ när $\mathcal{R}_0 > 1$.

Tyvärr gick Bienaymé's artikel och de få sidorna i Cournots bok helt obemärkt förbi vid den tiden. Artikeln uppmärksammades först på 1970-talet och boksidorna ytterligare tjugo år senare! Under tiden hade problemet och dess lösning återupptäckts av andra och ämnet hade utvecklats avsevärt. Vi återkommer till detta i kapitlen 9, 17 och 18.

Bienaymé var tvungen att lämna sitt arbete på finansministeriet efter 1848 års revolution. Professuren i sannolikhetsteori vid universitetet i Paris, som han säkert var den bästa kandidaten för, gavs också till någon annan. Trots detta kunde Bienaymé återigen arbeta för finansministeriet efter 1850, men han avgick 1852. Senare samma år valdes han in i vetenskapsakademien där han var specialist på statistikområdet. År 1853 bevisade han det som vissa moderna läroböcker kallar Bienaymé-Chebyshev-olikheten. År 1875 blev han ordförande för det nybildade *Société Mathématique de France*. Han dog i Paris 1878.

Ytterligare läsning

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. Brun, J., Robinet, A. (éd.): *A. Cournot, études pour le centenaire de sa mort*. Economica / Vrin, Paris (1978)
5. Cournot, A.-A.: *De l'origine et des limites de la correspondance entre l'algèbre et la géométrie*. Hachette, Paris (1847). archive.org
6. Doubleday, T.: *The True Law of Population* (1842). archive.org
7. Heyde, C.C., Seneta, E.: *I.J. Bienaymé*. Springer (1977)
8. Kendall, D.G.: The genealogy of genealogy: branching processes before (and after) 1873. *Bull. Lond. Math. Soc.* 7, 225–253 (1975)
9. Littré, É.: *Conservation, révolution et positivisme* (1852). gallica.bnf.fr
10. Malthus, T.R.: *An Essay on the Principle of Population* (1803). archive.org
11. Martin, T.: Antoine Augustin Cournot. In: Heyde, C.C., Seneta, E. (eds.) *Statisticians of the Centuries*, 152–156. Springer (2001)
12. Seneta, E.: Irenée-Jules Bienaymé. In: *ibid.*, 132–136.

Kapitel 8

Mendel och ärftlighet (1865)

År 1865 offentliggjorde Mendel resultaten av sina banbrytande experiment med korsningar av ärtor. Hans analys tillämpade grundläggande sannolikhetslära. Han betraktade även en dynamisk modell för en population av självpollinerande växter. Hans arbeten, som återfanns först år 1900, blev genetikens genombrott.

Johann Mendel föddes 1822 i Mähren, vid den tidpunkten en del av det österrikiska imperiet och nu är en del av Tjeckien. Hans far var bonde. Med sina goda resultat i gymnasiet och sin dåliga hälsa föredrog Mendel att fortsätta studera i stället för att arbeta på familjens gård. Men han hade inte råd att studera vid ett universitet. Men 1843 antogs han vid Saint Thomas-klostret i Brünn (nuvarande Brno), där han tog namnet Gregor. Han studerade teologi vid sidan av jordbruk. Han prästvigdes 1847. Han undervisade på en gymnasieskola i några år men misslyckades med ett prov för fast anställning. Mellan 1851 och 1853 kunde han tack vare stöd från andra som kände honom ändå fortsätta sina studier vid Wiens universitet där han läste fysik, matematik och naturvetenskap. Han återvände därefter till Brünn och undervisade i fysik vid en teknisk skola.



Figur 8.1:
Mendel (1822–1884)

Mellan 1856 och 1863 utförde Mendel en rad experiment på ett stort antal växter i klostrets trädgård. Han var medlem i det Naturhistoriska

sällskapet i Brünn och 1865 presenterade han sina resultat vid två av sällskapets möten. Hans arbete, *Försök med växtkorsningar*, publicerades på tyska året därpå i mötesprotokollet. Mendel förklarade hur han hade börjat studera variationerna hos ärter. Dessa växter förökar sig naturligt genom självpollinering och deras frön kan anta olika lätt identifierbara former: runda eller skrynkliga, gula eller gröna osv. När han korsade en växt som kommer från en stam med runda frön med en växt som kommer från en stam med skrynkliga frön, märkte han att deras korsningar alltid gav runda frön. Han kallade egenskapen ”runda frön” dominant och egenskapen ”skrynkliga frön” recessiv. På samma sätt visade han att egenskapen ”gula frön” var dominant medan egenskapen ”gröna frön” var recessiv.

Mendel lade sedan märke till att självpollinering av växter som odlats med hjälp av korsningar i den första generationen gav nya frön som hade antingen den dominerande eller den recessiva egenskaper i till synes slumpmässiga proportioner. Han märkte dessutom att när han upprepade experimentet många gånger i genomsnitt fick ungefär tre gånger fler frön med den dominanta egenskapen än med den recessiva egenskapen. Till exempel fick han i ett första experiment totalt 5 474 runda frön och 1 850 skrynkliga frön, vilket motsvarar ett förhållande på 2,96 till 1. Nästa experiment gav totalt 6 022 gula frön och 2 001 gröna frön, vilket motsvarar ett förhållande¹ på 3,01 till 1.

Mendel lade också märke till att bland de plantor som odlades från frön från den första generationen med den dominanta egenskapen var de som genom självpollinering gav frön med antingen den dominanta eller den recessiva egenskapen ungefär dubbelt så många som de som gav frön med enbart den dominanta egenskapen. Till exempel, bland de 565 plantor som odlades från runda frön från den första generationen gav 372 både runda och skrynkliga frön medan 193 gav endast runda frön. Förhållandet är lika med 1,93. På samma sätt gav 353 av de 519 plantor som odlades från gula frön i första generationen både gula och gröna frön medan 166 endast gav gula frön. Förhållandet är 2,13.

För att förklara dessa resultat fick Mendel den fantastiska idén att betrakta ett frös synliga egenskaper som ett resultat av två osynliga faktorer. Dessa faktorer var endera dominant (vi skriver *A*) eller recessiv (vi skriver *a*). Det finns sålunda tre möjliga kombinationer: *AA*, *Aa* och *aa*. Kombinationerna *AA* eller *Aa* tar sig uttryck i den dominant synliga egenskap som bärs

¹R. A. Fisher (se kapitel 14) konstaterade senare att sannolikheten för att få ett experimentellt resultat som ligger så nära det teoretiska värdet är mycket liten. Mendel arrangerade därmed troligen sina data. I det andra experimentet med $n = 6022 + 2001 = 8023$ frön så uppgår sannolikheten för att förhållandet ska skilja sig från 3 med mindre än 0,01 till bara cirka 10 %.

av A . Kombinationen aa tar sig uttryck i den recessiva synliga egenskapen a . Mendel antog dessutom att ståndarna (med pollen) och pistillerna (med äggceller) vid pollineringen kunde vara bärare av endast en av de två faktorerna, var och en med en sannolikhet $1/2$.

Korsning av homozygoterna AA och aa ger alltså heterozygotiska korsningar Aa med den dominanta synliga egenskapen A . Könscellerna hos korsningen Aa överför faktorn A med sannolikhet $1/2$ och faktorn a med sannolikhet $1/2$. Självpollinering av en växt som odlats från korsningen Aa ger därför AA med sannolikhet $1/4$, Aa med sannolikhet $1/2$ och aa med sannolikhet $1/4$. Se tabell 8.1.

Tabell 8.1: Resultattabell vid självpollinering av korsningen Aa . Sannolikheterna ges här som en funktion av de faktorer som överförs via ståndarna (i raderna) respektive pistillerna (i kolonnerna).

Faktor (Sannolikhet)	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 lade märke till att förhållandet $AA : Aa : aa$, som var $1 : 2 : 1$, de facto följde av kvadreringsregeln $(A + a)^2 = AA + 2Aa + aa$. Eftersom fröna AA och Aa har synliga egenskapen A medan endast fröna aa har den synliga egenskapen a , finns det faktiskt tre gånger fler frön med egenskapen A än med egenskapen a . Dessutom finns det i genomsnitt dubbelt så många frön Aa som AA . Självpollinering av växter som odlas från de förstnämnda ger frön med antingen den dominanta egenskapen (AA eller Aa) eller den recessiva egenskapen (aa). När det gäller självpollinering av växter som odlas från frön av typ AA så får genomgående frön AA med den dominanta egenskapen. Därmed förklaras alla observationer.

Mendel forstättade med att observera de påföljande generationerna. Om man utgick från N korsade frön Aa och för enkelhetens skull antog att varje planta genom självpollinering ger endast fyra nya frön, förutsade han att det genomsnittliga antalet frön $(AA)_n$, $(Aa)_n$ och $(aa)_n$ i generation n skulle beskrivas av tabell 8.2. För tydlighetens skull har delar vi resultaten i tabellen med N .

Dessa resultat erhålls enkelt med hjälp av de rekursiva sambanden

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

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

$$(aa)_{n+1} = (Aa)_n + 4(aa)_n, \quad (8.3)$$

Tabell 8.2: På varandra följande generationer.

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
totalt	1	4	16	64	256	1024

som säger att AA efter självpollinering ger fyra frön AA , att aa ger fyra frön aa och att Aa i genomsnitt ger ett frö AA , två frön Aa och ett frö aa . Mendel lade dessutom märke till att $(AA)_n = (aa)_n = 2^{n-1}(2^n - 1)$ och $(Aa)_n = 2^n$.

Ekvationen (8.2) och initialvärdet $(Aa)_0 = 1$ ger att $(Aa)_n = 2^n$. Insättning i ekvation (8.1) ger att $(AA)_{n+1} = 4(AA)_n + 2^n$. Vi inser lätt att $(AA)_n = c2^n$ är en partikulärlösning när $c = -1/2$. Den allmänna lösningen till den "homogena" ekvationen $(AA)_{n+1} = 4(AA)_n$ är $(AA)_n = C4^n$. Addition av dessa två lösningar ger att $(AA)_n = C4^n - 2^{n-1}$ uppfyller initialvärdet $(AA)_0 = 0$ om $C = 1/2$. Talföljden $(aa)_n$ uppfyller samma rekursionsformel och samma initialvärde som $(AA)_n$. Därmed är $(aa)_n = (AA)_n$.

Sammanfattningsvis delas andelen korsningar Aa i den totala populationen som ges av $2^n/4^n = 1/2^n$, med två i varje generation i självpollineringsfallet.

Ingen lade märke till Mendels arbeten under hans livstid. Några år senare upprepade Mendel även liknande experiment med andra växtarter. Han offentliggjorde artiklar om meteorologi och undersökte binas ärtflighet. Efter att ha blivit abbot 1868 ägnade han större delen av resten av sitt liv till att hantera administrativa problem. Han dog 1884.

Mendels arbete återupptäcktes till sist år 1900 nästan samtidigt av Hugo De Vries i Amsterdam, Carl Correns i Tübingen och Erich von Tschermak i Wien oberoende av varandra. Detta kom att inleda en ny era inom det som vi nu kallar genetik.

Ytterligare läsning

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

Kapitel 9

Galton och Watson och frågan om utdöende släktnamn (1873–1875)

År 1873 undersökte den brittiske statistikern Galton och hans landsman Watson problemet med utdöende släktnamn utan kännedom om Bienaymés arbete. Watson upptäckte att en rekursiv beräkning av den genererande funktionen för sannolikhetsfördelningen av antalet män i varje generation var möjlig. Hans analys av sannolikheten för utdöende var däremot felaktig.

Francis Galton föddes 1822, samma år som Mendel, nära Birmingham i England. Han var yngst av sju barn. Hans far var en rik bankdirektör. Han var kusin till Charles Darwin genom sin mor. Galton började studera medicin 1838, först på ett sjukhus i Birmingham och senare i London. Sommaren 1840 gjorde han sin första långa resa genom Europa ända till Istanbul. Därefter studerade han under fyra år vid *Trinity College*, universitetet i Cambridge. Hans far efterlämnade en betydande förmögenhet vid sin död 1844. Galton gav upp tanken på att bli läkare. Han reste till Egypten, Sudan och Syrien. Under de följande åren höll han sig till fritidssysselsättningar och tillbringade sin tid med jakt, resor i ballonger och båtar eller försökte förbättra den elektriska telegrafén. År 1850 startade han en forskningsexpedition till sydvästra Afrika (nuvarande Namibia). När han återvände till England 1852 valdes han in i *Royal Geographical Society*. Där kunde han följa nyheterna från expeditionerna till Östafrika för att leta efter Nilens källa. Han bosatte sig i London och skrev en mycket populär guide för resenärer. År 1856 valdes han in i *Royal Society*. Han var då intresserad av meteorologi och uppfann ordet ”anticyklon”. Efter att hans kusin Darwin 1859 publicerade *Om arternas uppkomst* intresserade sig Galton för ärftlighet. Han publicerade 1869 *Ärftligt geni*, där han hävdade att intellektuella förmågor kunde gå i arv.

År 1873 gav Alphonse de Candolle, en schweizisk botaniker, ut en bok med titeln *Historia om vetenskapen och vetenskapsmännen under de senaste två seklen*. Den innehöll även uppsatsen *Inverkan av arv, variation och urval för den mänskliga artens utveckling och en sannolik framtid för vår art*. Där gjorde han följande anmärkningar:



Figur 9.1: Galton (till vänster) och Watson (till höger).

”Bland den information och de åsikter som Benoiston de Châteauneuf, Galton och andra statistiker framförde, så såg jag inte den anmärkning som de borde ha gjort om det oundvikliga utdöendet av familjenamn. Det finns naturliga orsaker till att varje släktnamn dör ut [...] En matematiker borde kunna beräkna hur minskningen av namn uppför sig, utifrån kännedom om sannolikheten att få döttrar respektive söner och sannolikheten att ett givet par förblir barnlöst.”

Bienaymé studerade samma problem 1845. Candolle trodde att alla släktnamn var dömda att dö ut utan att känna till Bienaymé’s arbete. Galton lade märke till ovanstående stycke i Candolles bok. Galton formulerade det som ett öppet problem för läsarna av tidskriften *Educational Times*, även han, utan kännedom om Bienaymé’s arbete:

”Problem 4001: Låt oss betrakta en stor nation där vi endast tar hänsyn till antalet vuxna män, N , och de som bär olika släktnamn. Vi låter dem kolonisera ett område. Deras populationsdynamik beskrivs av att i varje generation har a_0 procent av de vuxna männen inga söner som når vuxen ålder; a_1 har en son; a_2 har två; och så vidare upp till a_5 som har fem.

Bestäm (1) den andel av deras efternamn som kommer att ha dött ut efter r generationer, och (2) i hur många fall släktnamnet kommer innehas av m personer.”

Märk att Bienaymé aldrig hade lyft den andra delen av problemet. Galton fick

inget tillfredsställande svar från tidskriftens läsare och kunde uppenbarligen inte själv finna lösningen på problemet. Han bad därför sin vän Henry William Watson, en matematiker, att ta sig an problemet.

Watson föddes i London 1827. Hans far var officer i den brittiska flottan. Han studerade först vid *King's College* i London och övergick därefter till matematik vid *Trinity College*, universitetet i Cambridge, från 1846 till 1850, bara några år efter Galton. Han gick därefter vidare i ämnet som stipendiat vid *Trinity College*, biträdande lärare vid *City of London School*, lektor i matematik vid *King's College* och slutligen professor i matematik vid *Harrow School* mellan 1857 och 1865. Han var förtjust i bergsbestigning och deltog i en expedition som nådde Monte Rosas topp i Schweiz 1855. Han prästvigdes till diakon 1856 och till anglikansk präst två år senare. Från 1865 till sin pensionering var han rektor i *Berkswell with Barton* nära Coventry, en tjänst som gav honom ymnigt med forskningstid.

Galton och Watson skrev tillsammans en artikel med titeln *Om sannolikheten för utdöende av släktnamn*, som publicerades 1875 i *Journal of the Royal Anthropological Institute*. Galton presenterade problemet och Watson förklarade sina beräkningar och de slutsatser han hade kommit fram till. De antog att män har högst q söner, där p_k är sannolikheten att ha k söner ($k = 0, 1, 2, \dots, q$). Därmed är blir $p_k = a_k/100$ med Galtons ursprungliga beteckningar och $p_0 + p_1 + \dots + p_q = 1$. Tänk på en situation där generation 0 består av en enda man. Generation 1 består av s män med sannolikheten p_s . Med hjälp av en metod som hade införts av Abraham de Moivre och blivit välkänd vid det här laget, betraktade Watson den genererande funktionen

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

som kan förknippas med sannolikheterna p_0, \dots, p_q . Låt på samma sätt $f_n(x)$ vara det polynom för vilket koefficienten för x^s är sannolikheten för att ha s män i generation n med utgångspunkt från en man i generation 0. Då är $f_1(x) = f(x)$. Watson lade märke till att sambandet

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

möjliggör en rekursiv beräkning av $f_n(x)$.

Sätt $f_n(x) = p_{0,n} + p_{1,n}x + p_{2,n}x^2 + \dots + p_{q,n}x^{(q^n)}$. Det finns då högst q^n män i generation n . Antag att det i generation $n - 1$ finns s män nummerade från 1 till s och låt t_1, \dots, t_s stå för det antal söner som var och en av dem har. Därmed kommer sannolikheten för att det att finnas

t män i generation n att vara lika med

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

När $s = 0$ så måste man förstå att denna sannolikhet är lika med 1 om $t = 0$ och lika med 0 om $t \geq 1$. Därför är

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

Av detta följer att

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

Speciellt är sannolikheten x_n för att släktnamnet ska dö ut inom n generationer lika med $p_{0,n}$. Denna sannolikhet är $f_n(0)$. Som ett första exempel använde Watson

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

dvs. $q = 3$ och $p_0 = p_1 = p_2 = 1/3$. Han beräknade polynomen $f_n(x)$ för $n = 1, \dots, 4$ med hjälp av ekvation (9.2). Han fick bland annat följande resultat

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

och $f_2(0) = 13/27 \approx 0,481$. Beräkningarna av $f_n(x)$ för $n \geq 3$ blir fort väldigt komplicerade. Faktum är att de blir så komplicerade att Watson börjar göra misstag redan i fallet $n = 4$. Eftersom $x_5 = f_5(0) = f_4(f(0))$ medför att beräkningen av $f_5(x)$ kan undvikas, fick han följande förteckning över sannolikheter för utdöende $x_n = f_n(0)$:

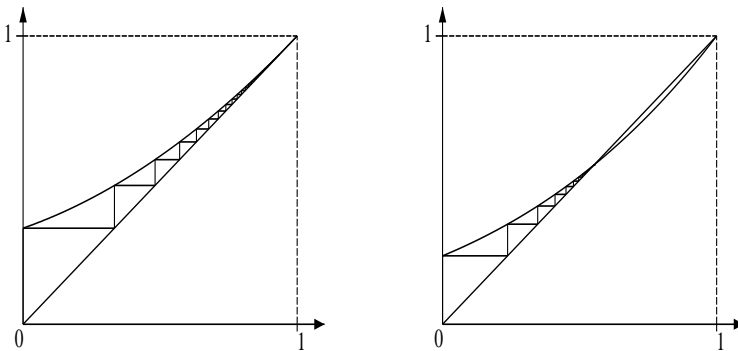
$$x_1 \approx 0,333, \quad x_2 \approx 0,481, \quad x_3 \approx 0,571, \quad x_4 \approx 0,641, \quad x_5 \approx 0,675.$$

De korrekta värdena kan verifieras med hjälp av Bienaymés samband $x_n = f(x_{n-1})$ och är $x_4 \approx 0,632$ och $x_5 \approx 0,677$. Den senare formeln kan även härledas från ekvation (9.2), se kapitel 17.

Watson märkte att varje man i genomsnitt har

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

söner och att $\mathcal{R}_0 = 1$ i hans första exempel. Man skulle alltså kunna tänka sig att om det ursprungliga antalet manliga familjemedlemmar var tillräckligt stort, skulle familjestorleken förbli ungefär konstant. Watson hävdade dock att utdöendesannolikheten x_n konvergerar mot 1 när $n \rightarrow +\infty$, om än ganska långsamt. Med andra ord skulle släktnamnet att dö ut, vilket Candolle hade föreslagit. Figur 9.2a, som inte återfinns i originalartikeln, och Bienaymés resultat bekräftar slutsatsen i det första exemplet.



Figur 9.2: Grafen för funktionerna $y = f(x)$ och $y = x$. Sannolikheten för utdöende $x_n = f(x_{n-1})$ inom n generationer är höjden på "trappsteget" n . Till vänster: $f(x) = (1 + x + x^2)/3$. Till höger: $f(x) = (3 + x)^5/4^5$.

Som ett andra exempel betraktade Watson binomialfördelningen

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

Den genererande funktionen (9.1) är nu

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

Han beräknade $f_2(x)$ och $x_2 = f_2(0)$. Då insåg han att $x_2 = f(x_1)$ och att $x_n = f(x_{n-1})$ för alla n . Men han trodde att denna formel var specifik för

binomialfallet (9.3). Genom att tillämpa den på fallet där $q = 5$, $a = 3$ och $b = 1$ fick han resultatet

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

Watson insåg att x_n måste konvergera mot en gräns x_∞ när $n \rightarrow +\infty$, som uppfyller

$$x_\infty = f(x_\infty) = \frac{(a + bx_\infty)^q}{(a + b)^q}.$$

Han noterade att $x = 1$ är en lösning på denna ekvation men insåg inte att det kunde finnas andra lösningar när $\mathcal{R}_0 > 1$. Så han drog felaktigt slutsatsen, vilseledd av Candolle, att utdöende alltid följer ($x_\infty = 1$) och att det även gäller det nyligen betraktade numeriska exemplet. Figur 9.2b, visar att så inte är fallet.

Watson lade märke till att medelantalet söner i detta exempel var större än 1 (man kan visa att $\mathcal{R}_0 = qb/(a + b) = 5/4$), vilket betyder att populationen tenderar att öka exponentiellt. Men detta hjälpte honom inte att upptäcka felet. Han spekulerade till och med i att släktnamnens utdöende var säkert för varje sannolikhetsfördelning (p_k), dvs. inte bara för binomialfallet. Vi återkommer till detta problem i kapitlen 17 och 18.

Galton fortsatte sin statistiska studie av familjer i boken *Engelska vetenskapsmän, deras natur och uppfostran*. Den fokuserade på *Royal Society* medlemmarnas stamtavla. Han intresserade sig även för mätningar av människokroppen. Han samlade vid en internationell utställning 1884 i London in uppgifter om ett stort antal människor. Hans resultat publicerades 1889 i en bok med titeln *Naturlig ärfilighet*. En bilaga återgav den artikel som han skrev i samarbete med Watson. I denna bok introducerades också en del ny statistisk vokabulär som "percentil" och "kvartil" samt ordet "arvshygien", dvs. förbättring av den mänskliga arten med utgångspunkt i ärftliga egenskaper. Efter 1888 utvecklade Galton en teknik för att känna igen fingeravtryck som några år senare skulle tas i bruk av den brittiska polisen. Han fortsatte även att studera vilken roll arv och miljö spelar för utvecklingen av fysiska och intellektuella egenskaper hos tvillingar, för storleken på årtor som odlats under flera generationer eller för färgen hos möss som avlats i ett laboratorium. Detta ledde honom till begreppet korrelation. År 1904 grundades Galtonlaboratoriet vid *University College* i London. Galton blev adlad 1909 och dog 1911.

Watson var författare till flera böcker. Särskilt noterbart är en avhandling om gasers kinetiska teori 1876 och en avhandling om den matematiska teorin om elektricitet och magnetism i två volymer (1885 och 1889). Han valdes in i *Royal Society* 1881 och dog i Brighton 1903.

Karl Pearson sammanfattade i andra delen av sin biografi om Galton 1924 artikeln om utdöende släktnamn utan att lägga märke till felet. Felet skulle slutligen uppmärksammas 1930 (se kapitel 18).

Ytterligare läsning

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

Kapitel 10

Lotka och teorin om stabila populationer (1907–1911)

År 1907 började den amerikanske kemisten Alfred Lotka studera sambandet mellan födelsetal, åldersspecifika dödstal och befolkningstillväxten med hjälp av en kontinuerlig tidsmodell. År 1911 publicerade han tillsammans med F. R. Sharpe en annan artikel om samma ämne, som även inkluderade åldersspecifika fertilitetstal. Den implicita ekvation som ger befolkningstillväxten kallas ofta "Lotkas ekvation".

Alfred James Lotka föddes av amerikanska föräldrar 1880 i Lemberg, som var en del av det österrikisk-ungerska imperiet (nu Lviv i Ukraina). Han studerade först i Frankrike och Tyskland och fick 1901 en kandidatexamen i fysik och kemi från universitetet i Birmingham i England. Därefter tillbringade han ett år i Leipzig där termodynamikens roll i kemi och biologi betonades av Wilhelm Ostwald, som skulle få Nobelpriset i kemi 1909. Lotka bosatte sig i New York 1902 och började arbeta för *General Chemical Company*.



Figur 10.1:
Lotka (1880–1949)

1907 och 1911¹ tog Lotka upp studiet av dynamiken hos ålders-

¹Den andra artikeln skrevs i samarbete med F. R. Sharpe, en matematiker från Cornelluniversitetet.

strukturerade befolkningar utan att känna till Eulers arbete i samma ämne (se kapitel 3). Till skillnad från Euler antog han att tid och ålder är kontinuerliga variabler. Låt $B(t)$ vara det manliga födelsetalet (antalet manliga födselar per tidsenhet) vid tiden t , $p(x)$ sannolikheten att fortfarande vara i livet vid ålder x och $h(x)$ fruktsamheten vid ålder x så att $h(x)dx$ är sannolikheten för en man att få en nyfödd son mellan ålder x och $x + dx$ om dx är försumbart liten. Då är $\int_0^{+\infty} p(x) dx$ den förväntade livslängden vid födseln. Dessutom är $B(t-x)p(x)dx$ antalet män som föds mellan tidpunkterna $t-x$ och $t-x+dx$ och som fortfarande lever vid tidpunkten t . Dessa män får $B(t-x)p(x)h(x)dx$ söner per tidsenhet vid tid t . Det totala antalet födda män vid tiden t är alltså

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

Lotka letade efter en exponentiell lösning för denna integralekvation i den okända $B(t)$ på formen $B(t) = be^{rt}$ och fick genom att dividera båda sidorna med $B(t)$ följande ekvation

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

som numera kallas "Lotkas ekvation" av demografer². Euler hade fått fram den analoga implicita ekvationen (3.1) för tillväxthastigheten när tid och ålder är diskreta variabler. Lotka noterade att den högra sidan av (10.1) är en avtagande funktion av r som går mot $+\infty$ när $r \rightarrow -\infty$ och som går mot 0 när $r \rightarrow +\infty$. Det finns alltså ett unikt värde för r , kalla det r^* , så att ekvation (10.1) håller. Dessutom är $r^* > 0$ om och endast om

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

Parametern \mathcal{R}_0 (beteckningen introducerades av Dublin och Lotka 1925) är det förväntade antalet söner som en man kan få under sitt liv.

Lotka menade att oavsett befolkningens ursprungliga åldersstruktur är antalet födda män per tidsenhet faktiskt sådant att $B(t) \sim be^{r^*t}$ när $t \rightarrow +\infty$, där b är en konstant³. Den totala befolkningen ges då av $P(t) = \int_0^{+\infty} B(t-x)p(x)dx$. Av detta följer att $P(t)$ också ökar eller minskar som e^{r^*t} när

²R. A. Fisher kom självständigt fram till samma ekvation 1927 och tolkade senare roten r^* som ett mått på "Darwinistisk lämplighet" i teorin om evolution genom naturligt urval.

³Detta bevisades rigoröst 1941 av William Feller, som då var professor i matematik vid *Brown University* i USA. En probabilistisk metod utvecklades 1968 av Crump, Mode och Jagers.

$t \rightarrow +\infty$: tillväxten är lika med r^* . Dessutom går befolkningens åldersstruktur, som ges av $B(t-x)p(x)/P(t)$, mot

$$\frac{e^{-r^*x}p(x)}{\int_0^{+\infty} e^{-r^*y}p(y)dy}.$$

Detta är vad Lotka kallade en ”stabil befolkning”: Ålderspyramiden behåller samma form över tid, men den totala befolkningen ökar eller minskar exponentiellt. Slutsatsen är alltså densamma som i Eulers diskreta tidsmodell, men i Lotkas studie tas hänsyn till fertilitetens åldersberoende. Den är alltså i någon mening mer allmän än Eulers.

Lotka fortsatte att arbeta med detta ämne under hela sitt liv. År 1908–1909 återupptog han sina studier vid Cornell för att avlägga en magisterexamen. Han arbetade för *National Bureau of Standards* från 1909 till 1911 och som redaktör för tidskriften *Scientific American Supplement* från 1911 till 1914. År 1912 erhöll han en doktorsexamen från universitetet i Birmingham genom att samla de artiklar han publicerat sedan 1907 om befolkningsdynamik och demografi. Under första världskriget arbetade han återigen för *General Chemical Company* med hur man kan fixera kväve från atmosfären. År 1920 gjorde en av hans artiklar om biologiska svängningar (se kapitel 13) ett djupt intryck på Raymond Pearl, en professor i biometri vid universitetet Johns Hopkins som just hade ”återupptäckt” den logistiska ekvationen (se kapitel 6). I hopp om att hitta ett jobb vid Rockefeller Institutet för medicinsk forskning i New York arbetade Lotka med de matematiska modeller som Ross utvecklat för malaria (se kapitel 12). Slutligen fick han ett tvåårigt stipendium från universitetet Johns Hopkins, vilket gjorde det möjligt för honom att skriva en bok med titeln *Element av fysikalisk biologi*, som publicerades 1925. Därefter blev han chef för forskningsavdelningen vid *Metropolitan Life Insurance Company* i New York. Han fokuserade på matematisk analys av demografiska frågor och publicerade flera böcker i samarbete med en kollega, statistikern och vice ordförande i företaget Louis Israel Dublin: *En människas penningvärde* (1930), *Livslängd* (1936) och *Tjugofem år av framsteg på hälsoområdet* (1937). Han valdes till ordförande för *Population Association of America* 1938–1939. Bland hans olika statistiska studier kan nämnas ”Lotkas lag” (som går tillbaka till 1926) som säger att antalet författare som har skrivit n artiklar inom ett visst vetenskapsområde minskar mer eller mindre som $1/n^2$ när n ökar.

Lotka publicerade också en bok på franska med titeln *Analytisk teori om biologiska associationer*. Den första delen, som var mer filosofisk, publicerades 1934. Den andra mer tekniska delen, som publicerades 1939, sammanfattade all hans forskning om mänsklig demografi sedan 1907. I sin

bok presenterade Lotka också sitt bidrag till problemet med utrotning av släktnamn. Efter publiceringen 1930 av Steffensens första artikel i ämnet (se kapitel 18) hade han tillämpat teorin på uppgifterna i 1920 års folkräkning av den vita befolkningen i USA. Han noterade att den observerade fördelningen $(p_k)_{k \geq 0}$ av antalet söner approximeras väl av en minskande geometrisk lag för alla $k \geq 1$: $p_0 = a$, $p_k = b c^{k-1}$ ($k \geq 1$), med $a = 0,4825$, $b = 0,2126$ och $c = 1 - b/(1 - a)$. På detta sätt blir $\sum_{k \geq 0} p_k = 1$. Den tillhörande genererande funktionen är $f(x) = a + b \sum_{k=1}^{+\infty} c^{k-1} x^k = a + \frac{bx}{1-cx}$. De två lösningarna till ekvationen $x = f(x)$ är $x = 1$ och $x = a/c$. Sannolikheten för utdöende x_∞ är den minsta av dessa två lösningar (se kapitel 7). Med de numeriska värdena för USA fann han $x_\infty \approx 0,819$, medan det genomsnittliga antalet söner var $\mathcal{R}_0 = f'(1) = (1-a)^2/b \approx 1,260$. Trots ett genomsnittligt antal barn (inklusive söner och döttrar) nära 2,5 är sannolikheten för att släktnamnet ska dö ut över 80 %.

Lotka valdes till ordförande för det amerikanska statistikförbundet 1942. Han gick i pension 1947 och dog 1949 i New Jersey. En ny upplaga av hans bok från 1925 utkom 1956 med den något annorlunda titeln *Element av matematisk biologi*.

Ytterligare läsning

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. Jagers, P.: Age-dependent branching processes allowing immigration. *Theor. Probab. Appl.* 13, 225–236 (1968).
7. Lotka, A.J.: Relation between birth rates and death rates. *Science* 26, 21–22 (1907) → Smith & Keyfitz (1977).
8. Lotka, A.J.: *Théorie analytique des associations biologiques*, 2^e partie. Hermann, Paris (1939) gallica.bnf.fr
9. Sharpe, F.R., Lotka, A.J.: A problem in age-distribution. *Philos. Mag. Ser. 6*, 21, 435–438 (1911) → Smith & Keyfitz (1977).
10. Smith, D.P., Keyfitz, N.: *Mathematical Demography*. Springer (1977)
11. Tanner, A.: *Von Molekülen, Parasiten und Menschen, Lotka und die Mathematisierung des Lebens*. Zürich (2014) doi:10.3929/ethz-a-010209129

Kapitel 11

Hardy-Weinbergs lag (1908)

År 1908 upptäckte den brittiske matematikern Hardy och den tyske läkaren Weinberg oberoende av varandra att i en oändligt stor population som parar sig slumpmässigt enligt Mendels lagar förblir frekvensen av genotyperna från två alleler konstant genom generationerna. Deras matematiska modell var en av utgångspunkterna för populationsgenetiken.

Godfrey Harold Hardy föddes 1877 i Surrey, England. Hans föräldrar var lärare. Han studerade matematik vid *Trinity College* vid universitetet i Cambridge från 1896, blev medlem av samma college 1900 och docent i matematik 1906. Efter sin första bok, *Integration av funktioner med en enda variabel* (1905), publicerade han 1908 *En kurs i ren matematik*, som omarbetades många gånger och översattes till många främmande språk.



Figur 11.1: Hardy (1877–1947)

Vid den tiden hade återupptäckten av Mendels arbete väckt vissa tvivel. Vissa biologer undrade varför de dominerande karaktärerna inte blev vanligare från generation till generation. Reginald Punnett, som hade skrivit en bok med titeln *Mendelism* 1905, ställde frågan till Hardy, som han spelade cricket med i Cambridge. Hardy skrev sin lösning i en artikel om *Mendelska proportioner i en blandad population*, som publicerades 1908. För att förenkla analysen föreställde han sig situationen med en stor population där valet av sexualpartner skulle vara slumpmässigt. Dessutom begränsade

han sin uppmärksamhet till endast två faktorer (eller ”alleler”) A och a , där A är dominant och a recessiv. För generation n , låt p_n vara frekvensen av ”genotypen” AA , $2q_n$ av Aa och r_n av aa . Naturligtvis är

$$p_n + 2q_n + r_n = 1.$$

Hardy antog också att ingen av dessa genotyper ledde till en överdödighet eller en minskad fertilitet jämfört med de två andra genotyperna. Frekvenserna i generation $n + 1$ kan enkelt beräknas genom att notera att en slumpmässigt vald individ i generation n överför allel A med en sannolikhet $p_n + q_n$: Antingen är genotypen AA och allelen A överförs med säkerhet eller så är genotypen Aa och allelen A överförs med 50% chans. På samma sätt överförs allelen a med en sannolikhet på $q_n + r_n$. Man kan därför konstruera tabell 11.1 på samma sätt som tabell 8.1.

Tabell 11.1: Beräkning av frekvensen av genotyperna i generation $n + 1$ från frekvensen av föräldrarnas alleler (raderna är för modern, kolumnerna för fadern).

Allel Frekvens	A $p_n + q_n$	a $q_n + r_n$
A	AA	Aa
$p_n + q_n$	$(p_n + q_n)^2$	$(p_n + q_n)(q_n + r_n)$
a	Aa	aa
$q_n + r_n$	$(p_n + q_n)(q_n + r_n)$	$(q_n + r_n)^2$

Frekvenserna av genotyperna AA , Aa och aa i generation $n + 1$ är p_{n+1} , $2q_{n+1}$ respektive r_{n+1} . Hardy fann alltså att

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

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

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

Han undersökte sedan under vilka förhållanden genotypens frekvens kunde förbli konstant genom generationerna och vara lika med p , $2q$ och r . Eftersom det per definition är $p + 2q + r = 1$ ser vi att ekvationerna (11.1)–(11.3) alla ger samma villkor $q^2 = pr$.

Den första ekvationen ger till exempel $p = (p + q)^2 = p^2 + 2pq + q^2$, vilket motsvarar $p(1 - p - 2q) = q^2$ och slutligen $pr = q^2$.

Med utgångspunkt från godtyckliga startvillkor $(p_0, 2q_0, r_0)$ med $p_0 + 2q_0 + r_0 = 1$, märkte Hardy att

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

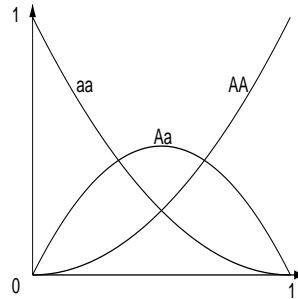
Tillståndet $(p_1, 2q_1, r_1)$ är därför redan en jämvikt. Så $(p_n, 2q_n, r_n)$ förblir lika med $(p_1, 2q_1, r_1)$ för alla $n \geq 1$. Om vi fastställer

$$x = p_0 + q_0$$

för frekvensen av allelen A i generation 0, så är $1 - x = q_0 + r_0$ frekvensen av allelen a . Genom att använda systemet (11.1)–(11.3) ännu en gång får vi följande resultat

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

för alla $n \geq 1$ (figur 11.2).



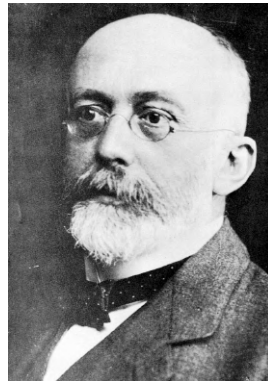
Figur 11.2: Grafer för funktionerna x^2 , $2x(1-x)$ och $(1-x)^2$ som motsvarar jämviktsfrekvenserna för genotyperna AA , Aa och aa .

Sammanfattningsvis leder ovanstående hypoteser till den lag enligt vilken frekvenserna av genotyperna AA , Aa och aa förblir oförändrade genom generationerna. Mendels teori leder inte till en progressiv ökning av frekvensen av den dominerande karaktären som man först trodde.

Några år senare skulle Fisher insistera på en viktig följd av denna lag: I en första approximation (dvs. om man antar att modellens hypoteser är realistiska) har en population en konstant genetisk varians. Denna iakttagelse löser ett av de problem som Darwins teori om evolution genom naturligt urval har gett upphov till. Darwin trodde nämligen, i likhet med sina samtida, att barnens fysiologiska egenskaper vid varje generation var ett slags genomsnitt av de två föräldrarnas egenskaper, där varje förälder bidrog med hälften. Denna idé har senare grundligt studerats med hjälp av statistik av Francis Galton och hans efterträdare vid biometrilaboratoriet, Karl Pearson. Om den var sann skulle variansen i dessa egenskaper i en population delas med två i varje generation och det skulle snart finnas en sådan homogenitet att

det naturliga urvalet, som antas förklara evolutionen, skulle vara omöjligt. Det skulle dock krävas flera år för att denna medelvärdesmekanism skulle förkastas, eftersom biometrikerna försvarade Darwins ståndpunkt och var ovilliga att erkänna att Mendels lagar är oundvikliga för att förstå evolutionen.

Efter detta arbete 1908 återvände Hardy till den rena matematiken. I sin självbiografi *En matematikers försvarstal* hävdar han till och med stolthet att han undvikit upptäckter av praktisk nytta. År 1910 valdes han in i *Royal Society*. År 1913 upptäckte han det indiska underbarnet Ramanujan och bjöd in honom att arbeta i Cambridge. Efter första världskriget blev han professor vid Oxfords universitet och fortsatte ett fruktbart samarbete med sin landsman Littlewood. Mellan 1931 och 1942 var han återigen professor i Cambridge. Han publicerade många böcker, ofta i samarbete: *Oändlighetens ordningsföljd* (1910), *Den allmänna teorin om Dirichlets serier* med Marcel Riesz (1915), *Ojämligheter* med Littlewood och Pólya (1934), *En introduktion till talteorin* med E. M. Wright (1938), *Ramanujan* (1940), *Fourier-serier* med Rogosinski (1944) och *Divergenta serier* (1949). Han dog i Cambridge 1947.



Figur 11.3:
Weinberg (1862–1937)

Flera decennier senare upptäckte man att Hardys lag för genfrekvenser också hade upptäckts samma år 1908 av den tyske läkaren Wilhelm Weinberg. Weinberg föddes i Stuttgart 1862. Efter att ha studerat i Tübingen och München fram till doktorsgraden i medicin hade han arbetat flera år på sjukhus i Berlin, Wien och Frankfurt. År 1889 bosatte han sig i Stuttgart som allmänläkare och förlösningsläkare. Trots att han var mycket upptagen med sitt arbete hade han funnit tid att skriva många artiklar i tyska vetenskapliga tidskrifter. År 1901 undersökte han statistiskt hur ofta tvillingar av samma kön förekommer. Artikelnen från 1908, i vilken han förklarade samma lag

som Hardy hade funnit, hade publicerats i en lokal vetenskaplig tidskrift och hade inte uppmärksammats. Men till skillnad från Hardy hade han fortsatt denna studie de följande åren och upptäckt till exempel generaliseringen till fallet där det finns mer än två alleler. Han hade också bidragit till området medicinsk statistik. Weinberg dog 1937. Efter återupptäckten av hans artikel från 1908 kallade genetiker lagen om genotypfrekvensernas stabilitet för "Hardy-Weinbergs lag".

Numera används denna lag ofta på följande sätt. Om en sällsynt recessiv allel a inte har något inflytande på överlevnad eller fertilitet och om vi känner till frekvensen x^2 av genotypen aa eftersom aa ger en viss fenotyp, kan vi beräkna x och uppskatta frekvensen $2x(1-x) \approx 2x$ av genotypen Aa . Om t. ex. frekvensen av aa är $1/20\,000$ får vi $x \approx 1/140$. Så $2x \approx 1/70$ är frekvensen av genotypen Aa . Den recessiva allelen a , som kan tyckas vara mycket sällsynt om man tittar på fenotyperna, är i själva verket inte så ovanlig.

Ytterligare läsning

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

Kapitel 12

Ross och malaria (1911)

År 1911 studerade den brittiske läkaren Ronald Ross, som redan hade fått Nobelpriset 1902 för sitt arbete om malaria, ett system av differentialekvationer som modellerade spridningen av denna sjukdom. Han visade att malaria kan bestå endast om antalet myggor ligger över en viss tröskel. Därför är det inte nödvändigt att döda alla myggor för att utrota malaria – det räcker att döda bara en viss del av dem. Liknande epidemiologiska modeller utvecklades senare av Kermack och McKendrick.

Ronald Ross föddes 1857 i norra Indien, där hans far var officer i den brittiska armén. Han studerade medicin i London men föredrog att skriva dikter och dramer. Efter att ha arbetat ett år på ett fartyg som kirurg lyckades han komma in i den indiska läkarkåren 1881. Hans medicinska arbete i Indien gav honom gott om fritid, under vilken han skrev litterära verk och lärde sig själv lite matematik. På permission till England 1888 tog han en examen i folkhälsa och studerade bakteriologi, en ny vetenskap som hade skapats några år tidigare av Pasteur och Koch. Tillbaka i Indien började Ross studera malaria. Under sin andra permission 1894 träffade han i London Patrick Manson, en specialist på tropiska sjukdomar, som i mikroskop visade honom vad den franske militärläkaren Alphonse Laveran hade upptäckt 1880: Blodet hos malariapatienter innehåller parasiter. Manson föreslog att parasiterna kunde komma från myggor, eftersom han själv i Kina hade upptäckt parasiten som orsakade en annan tropisk sjukdom (filariasis) i dessa insekter. Han trodde dock att människor smittades av parasiten när de drack vatten som förorenats av myggorna. Från 1895 till 1898 fortsatte Ross sin forskning i Indien och testade Mansons idé. År 1897 upptäckte han i magen på en viss myggart som han inte hade studerat tidigare (*Anopheles*) några parasiter som liknade dem som Laveran observerat. Sedan hans överordnade skickat honom till Calcutta under en säsong då malariafallen var sällsynta, beslöt han att studera malaria hos burfåglar. Han hittade parasiten i anophelesmyggornas spottkörtlar och lyckades experimentellt smitta friska fåglar genom att låta myggorna bita dem: Detta bevisade att malaria överförs genom myggbett och inte genom att man får i sig förorenat vatten. År 1899 lämnade Ross

den indiska läkarkåren för att undervisa vid *Liverpool School of Tropical Medicine*, som hade skapats ett år tidigare. Han valdes in i *Royal Society* 1901 och fick 1902 Nobelpriset i fysiologi eller medicin för sitt arbete om malaria. Han reste till Afrika, till Mauritius och i Medelhavsområdet för att popularisera kampen mot myggor. Metoden var framgångsrik i Egypten längs Suezkanalen, längs Panamakanalen som höll på att byggas, på Kuba och i Malaysia. Den var mindre framgångsrik i vissa andra områden. Ross publicerade en *Rapport om förebyggande av malaria på Mauritius* 1908 och *Prevention av malaria* 1910.



Figur 12.1: Ross (1857–1932)

Trots att han bevisade vissa myggors roll i överföringen av malaria mötte Ross tveksamhet när han hävdade att malaria kunde utrotas genom att minska antalet myggor. I den andra upplagan av sin bok om *Prevention av malaria*, som publicerades 1911, försökte han bygga upp matematiska modeller för överföring av malaria för att stödja sitt påstående. En av hans modeller bestod av ett system med två differentialekvationer. Låt oss införa följande beteckningar:

- N : den totala mänskliga befolkningen i ett visst område;
- $I(t)$: antalet människor som är infekterade med malaria vid tidpunkt t ;
- n : den totala myggpopulationen (som antas vara konstant);
- $i(t)$: antalet myggor som smittas av malaria;
- b : myggornas bettfrekvens;

- p (respektive p'): sannolikheten för överföring av malaria från människa till mygga (respektive från mygga till människa) under ett bett;
- a : Hur snabbt människor återhämtar sig från malaria;
- m : dödlighet hos myggor.

Under ett litet tidsintervall dt biter varje infekterad mygga bdt människor, av vilka en andel som är lika med $\frac{N-I}{N}$ ännu inte är infekterad. Med hänsyn till överföringssannolikheten p' finns det $b p' i \frac{N-I}{N} dt$ nya infekterade människor. Under samma tidsintervall är antalet människor som tillfrisknar $aI dt$. Därav följer att

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

På samma sätt biter varje icke-infekterad mygga bdt människor, bland vilka en andel I/N redan är infekterade. Med hänsyn till överföringssannolikheten p finns det $b p (n-i) \frac{I}{N} dt$ nya infekterade myggor. Om man antar att infektionen inte påverkar dödligheten bland myggor är antalet myggor som dör $mi dt$. Så

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

Eftersom malaria existerar permanent i de flesta smittade länder, betraktade Ross bara de stabila tillstånden i sitt system av två ekvationer: Antalet infekterade människor $I(t)$ och antalet infekterade myggor $i(t)$ förblir konstanta under tiden ($dI/dt = 0$ och $di/dt = 0$). För det första finns det alltid ett stabilt tillstånd med $I = 0$ och $i = 0$, vilket motsvarar frånvaro av malaria. För det andra sökte Ross efter ett stabilt tillstånd med $I > 0$ och $i > 0$ och fann att

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

Om ekvationerna för det stabila tillståndet divideras med produkten $I \times i$ blir problemet ett linjärt system av två ekvationer med två okända $1/I$ och $1/i$,

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

Lösningen är lätt att få fram.

Man kan notera att $I > 0$ och $i > 0$ om antalet myggor ligger över ett kritiskt tröskelvärde:

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

I detta fall motsvarar det stabila tillståndet en situation där sjukdomen är endemisk, dvs. ständigt närvarande. Ross drog slutsatsen att om antalet myggor n minskas under den kritiska tröskeln n^* , är det enda återstående stabila tillståndet $I = 0$ och $i = 0$, vilket innebär att malaria bör försvinna. Det är alltså inte nödvändigt att utrota alla myggor för att utrota malaria. Det är just detta som Ross ville betona med sin modell.

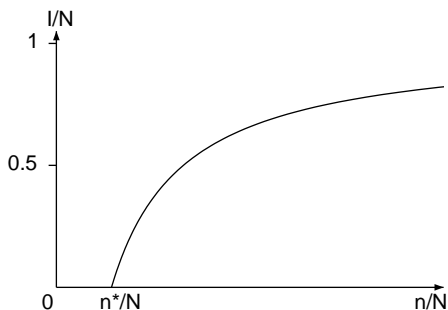
För att illustrera sin teori sökte Ross efter rimliga numeriska värden för parametrarna i sin modell. Han antog att

- dödligheten hos myggor är sådan att endast en tredjedel av dem lever efter tio dagar, så $e^{-10m} = \frac{1}{3}$ och $m = (\log 3)/10$ per dag;
- hälften av människorna är fortfarande smittade efter tre månader; så $e^{-90a} = 1/2$ och $a = (\log 2)/90$ per dag;
- en av åtta myggor biter varje dag, så $e^{-b} = 1 - 1/8$ och $b = \log(8/7)$ per dag.
- Infekterade myggor är vanligtvis inte smittsamma under de första tio dagarna efter smittan eftersom parasiterna måste genomgå flera omvandlingsstadier. Eftersom en tredjedel av myggorna kan överleva tio dagar antog Ross att ungefär en tredjedel av alla infekterade myggor är smittsamma: $p' = 1/3$;
- $p = 1/4$.

Ross kunde sedan med formeln (12.1) beräkna den infekterade fraktionen I/N i den mänskliga befolkningen som en funktion av förhållandet n/N mellan myggan och den mänskliga befolkningen. Han visade sina resultat i en tabell som motsvarar Fig. 12.2.

Kurvans form visar att andelen infekterade människor är högre än 50 % redan om förhållandet n/N ligger något över det kritiska värdet n^*/N . Men denna andel förändras inte mycket när förhållandet n/N ökar ytterligare. Detta förklarar varför korrelationen mellan antalet myggor och förekomsten av malaria aldrig tidigare hade uppmärksammats. Ross noterade dock att det numeriska värdet på tröskelvärdet n^*/N var mycket känsligt för små förändringar i bitfrekvensen b , men att detta inte ändrade den övergripande formen på kurvan i fig. 12.2. Hans kvalitativa förklaring är viktigare än

Figur 12.2: Andelen I/N av infekterade människor som en funktion av förhållandet n/N mellan myggan och den mänskliga populationen.



de kvantitativa resultaten, som lider av osäkerheten kring parametrarnas numeriska värden.

För att tolka det kritiska tröskelvärde n^* som Ross¹ upptäckte kan man tänka sig en infekterad människa som introduceras i en population av människor och myggor som båda är fria från malaria. Denna människa förblir i genomsnitt infekterad under en tidsperiod som är lika med $1/a$. Han eller hon får bn/N bett per tidsenhet, så i genomsnitt $bn/(aN)$ bett totalt under den tid han eller hon är infekterad. Han eller hon infekterar alltså i genomsnitt $bpn/(aN)$ myggor. Var och en av dessa infekterade myggor lever i genomsnitt under en tidsperiod som är lika med $1/m$, biter b/m människor och infekterar bp'/m människor. Efter överföringen från den första infekterade människan till myggorna och från dessa myggor till andra människor är det genomsnittliga antalet nyss infekterade människor produkten av de två föregående resultaten, dvs.

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

Detta \mathcal{R}_0 är antalet sekundära fall hos människor till följd av ett primärt fall hos människor. Infektionsprocessen, som sker kontinuerligt i tiden, kan alltså också betraktas som en diskret process med successiva generationer. Malaria kan "invadera" befolkningen endast om $\mathcal{R}_0 > 1$. Detta villkor är ekvivalent med $n > n^*$.

Sammanfattningsvis pläderade Ross mer allmänt för matematisk modellering inom epidemiologin:

¹Denna tolkning betonades först långt efter Ross arbete.

”I själva verket måste all epidemiologi, som handlar om sjukdomars variation från tid till annan eller från plats till plats, betraktas matematiskt, hur många variabler som än är inblandade, om den överhuvudtaget skall kunna betraktas vetenskapligt. Att säga att en sjukdom beror på vissa faktorer är inte mycket sagt, förrän vi också kan göra en uppskattning av hur mycket varje faktor påverkar hela resultatet. Och den matematiska metoden är egentligen inget annat än tillämpningen av noggranna resonemang på de aktuella problemen.”

Ross blev adlad 1911. Han flyttade till London och blev konsult för den brittiska armén under första världskriget. År 1923 publicerade han sin självbiografi, *Memoarer med en fullständig redogörelse för det stora malariaproblemet och dess lösning*. År 1926 invigdes *Ross Institute of Tropical Disease* (numera en del av *London School of Hygiene and Tropical Medicine*), som han blev direktör för. Ross dog i London 1932.

Ytterligare läsning

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)

Kapitel 13

Lotka, Volterra och rovdjurs- och bytesdjurs-systemet (1920–1926)

År 1920 studerade Alfred Lotka en rovdjurs- och bytesmodell och visade att populationerna kunde befinna sig i ständig svängning. Han utvecklade denna studie i sin bok *Element av fysikalisk biologi* från 1925. 1926 började den italienske matematikern Vito Volterra intressera sig för samma modell för att besvara en fråga som ställdes av biologen Umberto d’Ancona: Varför fångade fiskarna i Adriatiska havet fler rovfiskar under första världskriget, när fisketrycket var lågt?

År 1920 publicerade Lotka en artikel med titeln *Analytisk anmärkning om vissa rytmiska förhållanden i organiska system*. Sedan några år tillbaka hade han intresserat sig för vissa kemiska reaktioner som uppvisade märkliga övergående svängningar i laboratorieexperiment. Syftet med hans artikel var att antyda att ett system bestående av två biologiska arter till och med kunde svänga utan att svängningarna avtar. Det exempel han tog upp var en population av växtätare. I analogi med de ekvationer som används inom kemisk kinetik, låt $x(t)$ vara den totala massan av växter och $y(t)$ den totala massan av herbivorer vid tiden t . Lotka använde följande system av differentialekvationer som modell:

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

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

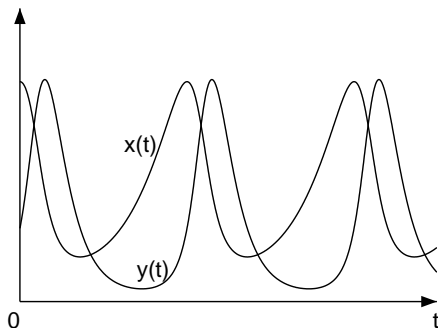
där parametrarna a , b , c och d alla är positiva. Parametern a är växternas tillväxttakt när det inte finns några växtätare, medan c är minskningstakten för populationen av växtätare när det inte finns några växter. Termerna $-bxy$ och dxy uttrycker att ju fler djur och växter det finns, desto större är massöverföringen från växter till djur (överföringen innefattar en viss massaförlust, så $d \leq b$). Genom att sätta $dx/dt = 0$ och $dy/dt = 0$ märkte Lotka att det finns två stabila tillstånd:

- $(x = 0, y = 0)$, populationen av växtätare är utdöd och det finns inga fler växter;

- $(x = c/d, y = a/b)$, växtätare och växter samexisterar.

Han skrev också utan bevis att om $(x(0), y(0))$ inte är ett av dessa två stabila tillstånd vid tiden $t = 0$, så oscillerar funktionerna $x(t)$ och $y(t)$ periodiskt: Det finns ett tal $T > 0$ som gör att $x(t+T) = x(t)$ och $y(t+T) = y(t)$ för alla $t > 0$ (figur 13.1)¹. Om växterna till exempel är mycket rikliga kommer populationen av växtätare att öka, vilket leder till en minskning av den totala mängden växter. När denna massa inte räcker till för att föda växtätarna dör vissa djur av hunger och den totala massan av växter börjar växa igen tills den når en nivå som motsvarar det ursprungliga värdet. Fenomenet upprepas således.

Figur 13.1:
Oscillationer i den totala massan av växter $x(t)$ och den totala massan av växtätare $y(t)$ som en funktion av tiden.



Lotka studerade modellen lite mer ingående i en andra artikel som publicerades 1920 med titeln *Odämpade svängningar som härrör från lagen om massverkan*. Han förklarade varför systemet kunde svänga på ett periodiskt sätt. Detta följer av det faktum att punkten $(x(t), y(t))$ måste hålla sig på en sluten bana i planet med x på den horisontella axeln och y på den vertikala axeln; närmare bestämt i den kvadrant där $x \geq 0$ och $y \geq 0$ (figur 13.2).

Genom att dividera ekvation (13.1) med ekvation (13.2) får vi efter en liten omskrivning följande resultat

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

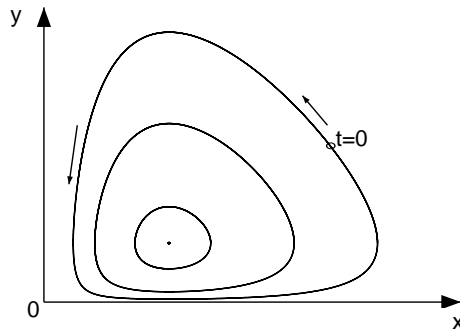
¹Perioden T beror på de initiala förhållandena, men Lotka insåg detta först 1925.

Integrationen ger

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

där K är en konstant som endast beror på det initiala tillståndet. Därför ligger punkten $(x(t), y(t))$ på kurvan $dx - c \log x = by - a \log y + K$, som råkar vara en sluten kurva (figur 13.2).

Figur 13.2: Diagram med växternas totala massa $x(t)$ på den horisontella axeln och växtätarnas totala massa $y(t)$ på den vertikala axeln. De tre slutna kurvorna runt det stabila tillståndet motsvarar olika begynnelsevillkor.



Banan för $(x(t), y(t))$ löper kring det stationära tillståndet $(c/d, a/b)$ moturs, vilket man lätt kan se genom att studera tecknet på dx/dt och dy/dt . Nära det stabila tillståndet uppvisar systemet små svängningar med en period som är lika med $2\pi/\sqrt{ac}$.

Låt $x = \frac{c}{d} + x^*$ och $y = \frac{a}{b} + y^*$, där $|x^*| \ll \frac{c}{d}$ och $|y^*| \ll \frac{a}{b}$. Då gäller

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

Från dessa två ekvationer får vi följande:

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

Dessa ekvationer är desamma som för den enkla pendelns svängningar i fysiken. Perioden är $2\pi/\sqrt{ac}$.

Raymond Pearl, som hade skickat den första artikeln från 1920 till *Proceedings of the National Academy of Sciences*, hjälpte Lotka att få ett tvåårigt stipendium från Johns Hopkins universitet för att skriva en bok med titeln *Element av fysikalisk biologi*. Boken publicerades 1925. I avsnittet som sammanfattar arbetet från 1920 nämns också att system med två arter, en värd- och en parasitart eller en bytes- och en rovdjursart, kan beskrivas med samma modell (13.1)–(13.2). Tyvärr väckte inte Lotkas bok någon större uppmärksamhet när den publicerades. Den berömda matematikern Volterra återupptäckte dock oberoende av Lotka samma modell kort därefter när han studerade ett fiskeproblem.

Vito Volterra föddes i det judiska gettot i Ancona 1860, strax före Italiens enande, när staden fortfarande tillhörde Kyrkostaten. Han hade inga syskon. Hans far, en tyghandlare, dog när Vito var två år gammal och lämnade familjen utan pengar. Volterra var en god elev i gymnasiet och lyckades trots fattigdomen fortsätta studera, först vid universitetet i Florens och senare vid *Scuola Normale Superiore* i Pisa. År 1882 doktorerade han i fysik och året därpå blev han professor i mekanik vid universitetet i Pisa. Han började arbeta vid universitetet i Turin 1892 och flyttade 1900 till en professur i matematisk fysik vid universitetet *La Sapienza* i Rom. Han blev senator 1905. Många av de föreläsningar han höll i Rom eller vid utländska universitet publicerades i bokform: *Tre lärdomar om några nya framsteg inom matematisk fysik* (Clark universitet, 1909), *Lektioner om integral- och integraldifferentialekvationer* (Rom, 1910), *Lektioner om linjefunktioner* (Paris, 1912), *Teorin om permutabla funktioner* (Princeton, 1912). Han tjänstgjorde som officer i den italienska armén under första världskriget och ledde byrån för krigsuppfinningar. Efter kriget deltog han aktivt i grundandet av den italienska matematiska unionen (1922) och det italienska nationella forskningsrådet (1923) och blev den sistnämndes första ordförande. Han blev också ordförande för Internationella kommissionen för vetenskaplig undersökning av Medelhavet (1923) och ordförande för *Accademia dei Lincei* (1924). En annan monografi, skriven i samarbete med J. Pérès, *Lektioner om sammansättning och permutabla funktioner*, publicerades 1924.

År 1925, vid 65 års ålder, blev Volterra intresserad av en studie av zoologen Umberto D'Ancona, som senare skulle bli hans svärson, om andelen broskfiskar (såsom hajar och rockor) som fiskades upp under åren 1905–1923 i tre hamnar i Adriatiska havet: Trieste, Fiume² och Venedig. D'Ancona hade

²Nu Rijeka i Kroatien.



Figur 13.3: Volterra (1860–1940) tilldelas doktorstitel *honoris causa* vid universitetet i Cambridge år 1900.

noterat att andelen av dessa fiskar hade ökat under första världskriget, då fisket hade minskat (tabell 13.1).

Tabell 13.1: Andelen upptagna broskfiskar i Trieste, Fiume och Venedig före, under och efter första världskriget.

År	1910	1911	1912	1913	1914	1915	1916
Trieste	5,7	8,8	9,5	15,7	14,6	7,6	16,2
Fiume	-	-	-	-	11,9	21,4	22,1
Venedig	21,8	-	-	-	-	-	-
År	1917	1918	1919	1920	1921	1922	1923
Trieste	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
Venedig	-	-	30,9	25,3	25,9	25,8	26,6

Eftersom broskfiskarna är rovdjur verkade det som om ett minskat fiske gynnade rovdjursarterna. Volterra, som inte kände till Lotkas arbete, förklarade denna observation med hjälp av samma modell,

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

där $x(t)$ står för antalet bytesdjur och $y(t)$ för antalet rovdjur. Han noterade, liksom Lotka, att detta system kan oscillera på ett periodiskt sätt med en

period T som beror på utgångstillståndet (x_0, y_0) . Han noterade också att

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

Genom att integrera över en period T (så att $x(0) = x(T)$ och $y(0) = y(T)$) fick han följande resultat

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

Genomsnittet under en period av både antalet bytesdjur och antalet rovdjur är alltså oberoende av de ursprungliga förhållandena. Om fisket minskar, ökar dessutom tillväxten a för bytesdjur medan utrotningstakten c för rovdjur minskar. Därför minskar genomsnittet för $x(t)$ medan genomsnittet för $y(t)$ ökar: Andelen rovdjur ökar. Detta är precis vad som observerats i statistiken från fisket i Adriatiska havet.

Volterra publicerade sin artikel först på italienska 1926. En engelsk sammanfattning publicerades några månader senare i *Nature*. Lotka informerade Volterra och andra vetenskapsmän om att hans studie av rovdjurs- och bytesdjurssystem publicerades före Volterras, men hans artikel från 1920 och hans bok från 1925 skulle inte alltid nämnas. Lotka arbetade då redan för ett försäkringsbolag, så hans arbete inriktades på mänsklig demografi. Volterra fortsatte att arbeta med varianter av rovdjurs-bytesystemet under ett decennium. Han höll en serie föreläsningar 1928–1929 vid det nybildade Institut Henri Poincaré i Paris. Anteckningarna till dessa föreläsningar publicerades 1931 under titeln *Lärdomar om den matematiska teorin om kampen för livet*. År 1935 publicerade Volterra i samarbete med Umberto D'Ancona ytterligare en bok om *Biologiska samband ur matematisk synvinkel*.

Även om rovdjurs- och bytesmodellen tycks förklara fiskedata korrekt, hade debatten om huruvida förenklade modeller är realistiska inom ekologin precis börjat och är fortfarande föremål för vetenskapliga tvister. Numera är rovdjurs- och bytesmodellen även känd som Lotka- och Volterra-modellen och är en av de mest använda inom ekologin.

År 1931 vägrade Volterra att ansluta sig till Mussolini. Han förlorade sin professur vid universitetet i Rom och uteslöts från de italienska vetenskapliga akademierna, där han var en av de mest kända medlemmarna. Från och med då höll han sig huvudsakligen utanför Italien, reste runt i Europa och höll föreläsningar. Han publicerade tillsammans med J. Pérès den första delen av en *Allmän teori om funktionaler* (1936) och en bok tillsammans med B. Hostinský om *Infinitesimala linjära operationer* (1938). Han dog i Rom 1940.

Ytterligare läsning

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](https://www.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](https://www.archive.org)
7. Lotka, A.J.: *Elements of Physical Biology*. Williams & Wilkins, Baltimore (1925) [archive.org](https://www.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](https://www.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](https://www.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)

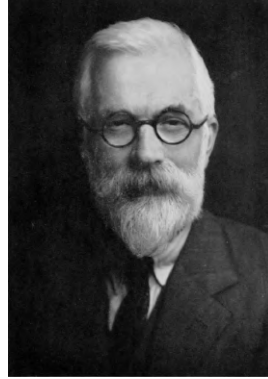
Kapitel 14

Fisher och det naturliga urvalet (1922)

1922 publicerade den brittiske matematikbiologen Ronald Fisher en mycket inflytelserik artikel om populationsgenetik. I detta kapitel behandlas endast ett avsnitt av artikeln, som fokuserar på en variant av Hardy-Weinberg-modellen som inkluderar naturligt urval. Fisher visade att om heterozygoten gynnas kan båda allelerna samexistera. Om en av de två homozygoterna gynnas försvinner den andra allelen. Det underliggande problemet är att förklara varför vissa gener kan ha flera alleler.

Ronald Aylmer Fisher föddes i London 1890 som det sista av sex barn. Hans far var auktionsförrättare, men gick senare i konkurs. Fisher studerade matematik och fysik vid *Gonville and Caius College* vid universitetet i Cambridge mellan 1909 och 1913. Genetiken utvecklades snabbt vid den tiden. Från och med 1911 deltog Fisher i mötena i *Eugenics Society* som initierades av Galton. Han började inrikta sig på statistiska problem med anknytning till Galtons och Mendels arbete. Efter att ha avslutat sina universitetsstudier tillbringade han en sommar med att arbeta på en gård i Kanada och arbetade sedan för *Mercantile and General Investment Company* i City av London. På grund av sin extrema närsynthet kunde han inte delta i första världskriget trots att han hade anmält sig frivilligt. Han tillbringade dessa år med att undervisa i gymnasieskolor. På fritiden tog han hand om en gård och fortsatte sin forskning. Han fick viktiga nya resultat som kopplade samman korrelationskoefficienter med mendelsk genetik. År 1919 började han arbeta som statistiker vid Rothamsted Experimenteringsstation, som var inriktad på jordbruk.

1922 publicerade Fisher en artikel med titeln *Om dominansförhållandet*. Bland flera andra viktiga nya idéer tog artikeln upp en matematisk modell som kombinerade Mendels lagar och den idé om naturligt urval som Darwin betonade i sin evolutionsteori. Fisher betraktade samma situation som Hardy med två alleler A och a och med hypotesen om slumpmässig parning. Men han antog att individer med genotyperna AA , Aa och aa har olika dödlighet innan de når vuxen ålder, vilket efterliknar det naturliga urvalet. Om man sätter p_n , $2q_n$ och r_n för frekvensen av de tre genotyperna bland vuxna individer i generation n , finns det $(p_n + q_n)^2$, $2(p_n + q_n)(q_n + r_n)$ respektive



Figur 14.1:
Fisher (1890–1962)

$(q_n + r_n)^2$ nyfödda barn i generation $n + 1$ som har dessa genotyper. Låt u , v och w vara respektive överlevnadssannolikheter från födelse till vuxen ålder. Frekvenserna av genotyperna bland vuxna individer i generation $n + 1$ är då p_{n+1} , $2q_{n+1}$ och r_{n+1} , med

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

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

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

där vi för enkelhetens skull låter

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

Om vi kommer ihåg att $p_n + 2q_n + r_n = 1$ ser vi att när $u = v = w$ (dvs. när det inte finns något naturligt urval) reduceras systemet (14.1)–(14.3) till det system (11.1)–(11.3) som Hardy hade studerat.

Låt

$$x_n = p_n + q_n$$

vara frekvensen av allelen A bland vuxna individer i generation n . Då är $q_n + r_n = 1 - x_n$ frekvensen av allelen a . Genom att addera (14.1) och (14.2) får vi följande resultat:

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

Denna ekvation kan skrivas om i följande form:

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

Det finns alltid minst två stabila tillstånd där frekvensen x_n förblir konstant över generationerna: $x = 0$ (populationen består helt av homozygota aa) och $x = 1$ (populationen består helt av homozygota AA).

Med hjälp av ekvation (14.4) kan man visa att om den homozygota AA har större chans att överleva än de två andra genotyperna ($u > v$ och $u > w$), kommer allelen a att försvinna successivt ur populationen. Detta fall borde inte vara särskilt vanligt i naturen om vi vet att båda allelerna samexisterar. Om däremot den heterozygota Aa har en selektiv fördel jämfört med de homozygota AA och aa ($v > u$ och $v > w$) kan de tre genotyperna samexistera i populationen. Detta är det vanligaste fallet och kan förklara den "livlighet" hos hybrider som uppmärksammats av jordbrukare.

Det stabila tillståndet $x = 1$ är stabilt när $u > v$ eftersom $x_{n+1} - x_n \approx (1 - x_n)(u - v)/u$ när x_n är nära 1. Populationen går mot detta stabila tillstånd. Det stabila tillståndet $x = 1$ är instabilt när $u < v$, i vilket fall det finns ett tredje stabilt tillstånd

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

med $0 < x^* < 1$. Dessutom kan vi kontrollera att detta är stabilt. Det stabila tillståndet x^* motsvarar en blandning av de tre genotyperna.

Genom att helt enkelt kombinera Mendels lagar och en hypotes om naturligt urval (här olika överlevnadssannolikheter för de tre genotyperna) kan vi förklara de två situationerna med samexistens eller försvinnande av genotyper. Efter Fisher utvecklades denna modell även av J. B. S. Haldane (se kapitel 17) och av Sewall Wright (se kapitel 19).

I väntan på kapitel 20, lägg märke till att om A är helt dominant och den homozygota aa är missgynnad jämfört med de två andra genotyperna, med siffrorna $u : v : w$ i förhållandet $1 : 1 : 1 - \varepsilon$, så blir ekvation (14.4)

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

för $\varepsilon \ll 1$. Om överlevnaden hos den heterozygota Aa ligger halvvägs mellan den hos de två homozygota, så är siffrorna $u : v : w$ i förhållandet $1 : 1 - \varepsilon/2 :$

$1 - \varepsilon$ och

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

när $\varepsilon \ll 1$.

På Rothamsted analyserade Fisher långtidsdata om skördar och meteorologi. Men han bidrog också i hög grad till den statistiska metodiken. År 1925 publicerade han en bok med titeln *Statistiska metoder för forskare*, som blev mycket framgångsrik och trycktes om många gånger. Han blev medlem av *Royal Society* 1929. År 1930 publicerade Fisher en bok om *Den genetiska teorin om naturligt urval*, en milstolpe i populationsgenetikens historia. Han blev professor i eugenik vid *University College* i London 1933 och efterträdde Karl Pearson vid Galtonlaboratoriet. År 1943 flyttade han till en professur i genetik vid universitetet i Cambridge, denna gång som efterträdare till R. C. Punnett (se kapitel 11). Han publicerade också flera böcker: *Försöksplanering* (1935), *Teorin om inavel* (1949) och *Statistiska metoder och vetenskaplig slutsats* (1956). Han blev adlad 1952, bosatte sig i Australien efter att ha gått i pension 1959 och dog i Adelaide 1962. Vi återkommer till en annan del av hans arbete i kapitel 20.

Ytterligare läsning

1. Fisher Box, J.: R.A. Fisher, *The Life of a Scientist*. John Wiley & Sons, New York (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)

Kapitel 15

Yule och evolution (1924)

1924 studerade den brittiske statistikern Yule en utvecklingsmodell där arter kan bilda nya arter genom små mutationer och släkten kan bilda nya släkten genom stora mutationer. Hans syfte var att förklara fördelningen av antalet arter inom släkten, där de flesta släkten innehåller endast en art och några få släkten innehåller ett stort antal arter. Den stokastiska "födelseprocess" som Yule införde i sin modell är fortfarande ett grundläggande verktyg vid studiet av fylogenetiska träd och på många andra områden.

George Udny Yule föddes i Skottland 1871 och hans far hade en hög position inom den brittiska administrationen i Indien. Vid 16 års ålder började Yule studera vid *University College* i London för att bli ingenjör. År 1892 ändrade han inriktning och tillbringade ett år med att forska i Bonn under överinseende av fysikern Heinrich Hertz, som några år tidigare hade visat att elektromagnetiska vågor existerade. När Yule återvände till England erbjöd Karl Pearson honom en tjänst som biträdande professor i tillämpad matematik vid *University College*. Yule började efter Pearson att inrikta sig på statistik. År 1911 publicerade han *En introduktion till teorin om statistik*, som trycktes om 14 gånger. Året därpå flyttade han till universitetet i Cambridge. Hans forskningsarbete handlade om teoretiska aspekter av statistiken men också om tillämpningar inom jordbruk och epidemiologi. Han blev ledamot av *Royal Society* 1922.

1924 publicerade Yule en artikel med titeln *En matematisk evolutionsteori baserad på Dr. J. C. Willis slutsatser*. Willis var en kollega från *Royal Society* som 1922 hade publicerat en bok med titeln *Ålder och territorium, en studie om geografisk spridning och arternas ursprung*. Han hade studerat fördelningen av arter mellan olika släkten i klassificeringen av växter och djur. De uppgifter som han hade sammanställt visade att de flesta släkten endast innehöll en art, att allt färre släkten innehöll ett större antal arter och att det fortfarande fanns ett fåtal släkten som innehöll ett stort antal arter. Tabell 15.1 visar uppgifterna om ormar, ödlor och två familjer av skalbaggar (*Chrysomelidae* och *Cerambycinae*).

De 1 580 arter av ödlor som var kända vid den tiden hade klassificerats i 259 släkten, 105 släkten som innehöll endast en art, 44 endast två arter, 23



Figur 15.1:
Yule (1871–1951)

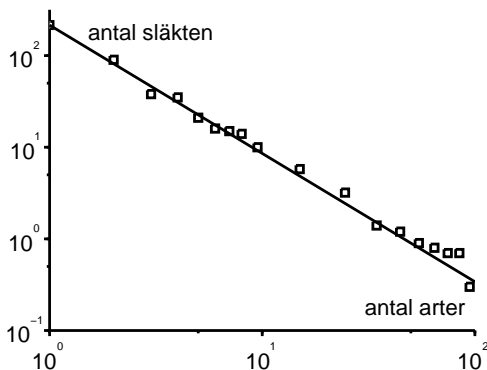
Tabell 15.1: Uppgifter sammanställda av Willis.

Antal arter	Antal släkten			
	<i>Chrysomelidae</i>	<i>Cerambycinae</i>	Ormar	Ödlor
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
totalt	627	1 024	293	259

endast tre arter osv. och två släkten som innehöll mer än hundra arter. För andra djur- och växtfamiljer hade fördelningen av släkten enligt antalet arter de innehåller en mycket liknande form.

Yule föreslog att Willis skulle försöka plotta sina data i ett diagram med logaritmiska skalor. Detta gav ett slående resultat (figur 15.2): Logaritmen av antalet Q_n släkten som innehåller n arter minskar mer eller mindre linjärt med $\log(n)$. Med andra ord finns det konstanter $\alpha > 0$ och $\beta > 0$ som gör att $Q_n \approx \alpha n^{-\beta}$; fördelningen följer en ”potenslag”. I sin artikel från 1924 sökte Yule efter en matematisk modell för evolutionen som kunde förklara en sådan statistisk fördelning.

Figur 15.2: Antalet släkten som en funktion av antalet arter som de innehåller, med decimallogaritmisk skala. Uppgifter för *Chrysomelidae*. För att jämna ut fluktuationerna när n (antalet arter) är stort räknades släkterna för intervall av n -värden enligt tabell 15.1. Det genomsnittliga antalet släkten för ett enda värde på n kan således vara mindre än 1.

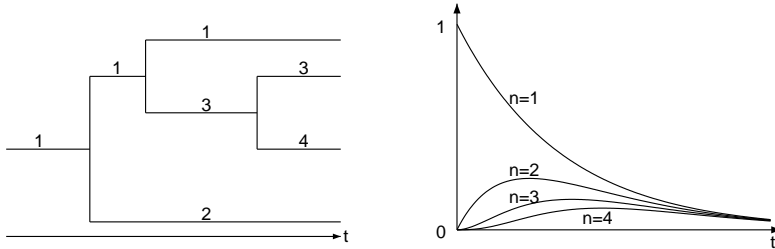


För detta ändamål tänkte han sig först en stokastisk modell för kontinuerlig tid¹ för tillväxten av antalet arter inom ett släkte (figur 15.3a). Han började med endast en art vid tiden $t = 0$ och antog att sannolikheten för att en art genom mutation skulle ge upphov till en ny art av samma släkte under ett ”litet” tidsintervall dt (på evolutionens tidsskala) var lika med $r dt$ med $r > 0$.

Låt $p_n(t)$ vara sannolikheten för att det finns n arter vid tiden t (n är ett heltal men t är ett reellt tal). För att beräkna $p_n(t + dt)$ har Yule övervägt flera fall:

- Om det finns $n - 1$ arter vid tiden t har varje art en sannolikhet $r dt$ att

¹McKendrick (se kapitel 16) hade redan börjat studera sådana modeller inom populationsdynamik i en artikel som publicerades 1914.



Figur 15.3: (a) En simulering av utvecklingen av antalet arter inom ett släkte. Art 1 ger upphov till art 2 och 3. Art 3 ger upphov till art 4. (b) Sannolikheten $p_n(t)$ att det finns n arter av samma genus vid tiden t , för $1 \leq n \leq 4$.

generera en ny art mellan t och $t + dt$; i gränsen $dt \rightarrow 0$ kommer det att finnas n arter vid tiden $t + dt$ med en sannolikhet $(n - 1) r dt$;

- Om det finns n arter vid tiden t kommer det att finnas $n + 1$ arter vid tiden $t + dt$ med en sannolikhet på $n r dt$.

Således ges $p_n(t)$ av följande system av differentialekvationer

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

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

för alla $n \geq 2$. Från den första ekvationen får vi $p_1(t) = e^{-rt}$ eftersom $p_1(0) = 1$. Det är möjligt att visa att lösningen till den andra ekvationen som uppfyller begynnelsevillkoret $p_n(0) = 0$ är

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

för alla $n \geq 2$ (figur 15.3b). Så vid en fast tidpunkt t är sannolikhetsfördelningen $(p_n(t))_{n \geq 1}$ ”geometrisk” med ett förhållande mellan två på varandra följande termer som är lika med $1 - e^{-rt}$.

Vi noterar först att ekvation (15.2) är ekvivalent med

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

från vilken vi successivt kan beräkna $p_2(t)$, $p_3(t)$... Vi får $p_2(t) = e^{-rt}(1 - e^{-rt})$, sedan $p_3(t) = e^{-rt}(1 - e^{-rt})^2$, vilket tyder på formel (15.3) för den allmänna lösningen. Man kan slutligen kontrollera att denna formel är en lösning av ekvation (15.4).

Yule drog också slutsatsen från formeln (15.3) att det förväntade antalet arter ökar exponentiellt med tiden: $\sum_{n=1}^{+\infty} n p_n(t) = e^{rt}$.

Vi noterar först att för $|x| < 1$,

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

Därefter

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

Om T är den fördubblingstid som definieras av $e^{rT} = 2$, är sannolikhetsfördelningen $(p_n(t))_{n \geq 1}$ av antalet arter vid tiden $t = T$ geometrisk med en kvot på $1/2$: $1/2, 1/4, 1/8, 1/16$... Vid tiden $t = kT$ är den geometrisk med förhållandet $1 - 1/2^k$ och $p_1(kT) = 1/2^k$.

Yule ansåg sedan att det parallellt med ökningen av antalet arter som tillhör samma släkte skulle ske en liknande process på grund av större mutationer som leder till skapandet av nya släkten. Låt $s dt$ vara sannolikheten för att ett befintligt släkte ska generera ett nytt släkte under ett litet tidsintervall dt . Liksom tidigare, om man antar att det bara finns ett släkte vid tiden $t = 0$, är det förväntade antalet släkten vid tiden t lika med e^{st} . Det genomsnittliga antalet släkten som skapas per tidsenhet vid tiden t är derivatan $s e^{st}$. I gränsen² där $t \rightarrow +\infty$ är det genomsnittliga antalet släkten som vid tiden t har existerat mellan x och $x + dx$ tidsenheter då $s e^{s(t-x)} dx$. Sannolikheten för att ett slumpmässigt valt släkte vid tiden t ska ha funnits mellan x och $x + dx$ tidsenheter är $s e^{-sx} dx$.

Om ett släkte som väljs slumpmässigt vid tiden t har existerat mellan x och $x + dx$ tidsenheter, är sannolikheten att detta släkte innehåller n arter, enligt formel (15.3), lika med $e^{-rx}(1 - e^{-rx})^{n-1}$ för alla $n \geq 1$. Sannolikheten q_n för att ett släkte som slumpmässigt väljs vid tiden t ska innehålla n arter är

²Yule tog också hänsyn till det fall där t inte kan antas vara särskilt stor jämfört med fördubblingstiden för e^{rt} . Beräkningarna är lite mer komplicerade, men slutresultaten skiljer sig inte nämnvärt från varandra.

alltså följande

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

Låt $u = r/s$. En enkel beräkning visar att $q_1 = 1/(1+u)$ och att

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

för alla $n \geq 2$.

Vi har faktiskt $(1 - e^{-rx})^{n-1} = (1 - e^{-rx})^{n-2} (1 - e^{-rx})$. Så

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

Genom partiell integration får vi

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

Formeln (15.5) visar att sekvensen av sannolikheter $(q_n)_{n \geq 1}$ är avtagande. Maximalvärdet uppnås alltså för $n = 1$: De flesta släkten innehåller endast en art. Detta är precis vad uppgifterna hade visat. Dessutom är minskningen av q_n mot 0 när n går mot oändligheten relativt långsam eftersom $q_n/q_{n-1} \rightarrow 1$. Detta kan förklara varför vissa släkten innehåller ett stort antal arter. Mer exakt visade Yule att $\log q_n$ minskar linjärt med $\log(n)$.

Introduktion av Eulers gammafunktion $\Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-t} dt$. Då gäller $\Gamma(n+1) = n! = n \times (n-1) \times \cdots \times 2 \times 1$ när n är ett heltal och $\Gamma(z+1) = z\Gamma(z)$. Så (15.5) tar formen

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

Men Stirlings approximation ger $\log \Gamma(n) \approx n \log n - n - \frac{1}{2} \log n +$ konstant. På samma sätt är $\log \Gamma(n+1+1/u) \approx n \log n - n + (\frac{1}{u} + \frac{1}{2}) \log n +$ konstant. Till sist: $\log q_n \approx -(1+\frac{1}{u}) \log n +$ konstant.

Tänk till exempel på ödlor. Parametern u kan uppskattas utifrån andelen $q_1 = 1/(1+u)$ av de släkten som endast innehåller en art. Enligt tabell 15.1

har vi $q_1 = 105/259$ så $u \approx 1,467$. Vi kan sedan beräkna den teoretiska sannolikheten q_n och det förväntade antalet Q_n släkten som innehåller n arter genom att multiplicera q_n med det totala antalet arter, vilket är 259 (tabell 15.2). Yule noterade att överensstämmelsen mellan observationerna och beräkningarna är relativt god³ med tanke på modellens enkelhet, som inte tar hänsyn till till till exempel de katastrofer som arterna har korsat genom miljontals år av evolution.

Tabell 15.2: Jämförelse mellan data och teori när det gäller ödlor (1580 arter klassificerade i 259 släkten).

Antal arter per släkte	Observerat antal släkten	Beräknat antal släkten
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
totalt	259	259

Efter 1931 drog sig Yule successivt tillbaka från universitetet i Cambridge. Han började intressera sig för den statistiska fördelningen av meningslängden för att identifiera bokförfattare. Han tillämpade detta särskilt

³För antalet släkten som innehåller mer än 100 arter fick Yule en bättre överensstämmelse än i tabell 15.2 genom att beakta att t inte var stor jämfört med fördubblingstiden e^{st} .

på den bok som publicerades av John Graunt (se kapitel 2) men möjligen inspirerad av William Petty. År 1944 publicerade han en bok om *Statistisk undersökning av litterärt ordförråd*. Han dog 1951.

Numera används Yules modell fortfarande för att analysera ”fylogenetiska träd” (arternas genealogiska träd). Dessa träd, som liknar träden i figur 15.3, är mer kända tack vare de nya uppgifterna från molekylärbiologin. Men tillämpningarna av den stokastiska process som definieras av ekvationerna (15.1)–(15.2) är inte begränsade till evolutionsteorin. Denna process är en byggsten i många modeller inom populationsdynamik, från den mikroskopiska nivån (för att modellera till exempel kolonier av bakterier) till den makroskopiska nivån (för att modellera början av en epidemi). Den kallas ”ren födelseprocess” eller ”Yule-process”. En enkel variant inkluderar en sannolikhet $m dt$ att dö under ett litet tidsintervall dt : Den förväntade populationsstorleken vid tiden t för denna ”födelse- och dödsprocess” är då $e^{(r-m)t}$. När det gäller sannolikhetsfördelningen (15.5) kallas den ibland för Yule-fördelningen. Fördelningar med svansar som uppfyller potenslagar har väckt stor uppmärksamhet inom olika vetenskapsområden. Studiet av epidemier i slumpmässiga nätverk fördelade enligt en potenslag är bara ett exempel.

Ytterligare läsning

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

Kapitel 16

McKendrick och Kermack om modellering av epidemier (1926–1927)

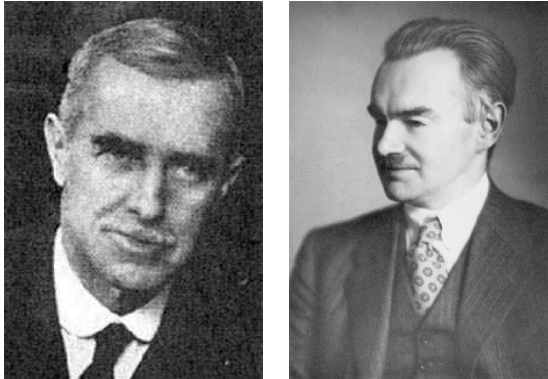
1926 studerade McKendrick en stokastisk epidemimodell och fann en metod för att beräkna sannolikheten för att en epidemi ska nå en viss storlek. Han upptäckte också den partiella differentialekvation som styr åldersstrukturerade populationer. År 1927 studerade Kermack och McKendrick en deterministisk epidemimodell och fick fram en ekvation för den slutliga epidemins storleken, en modell som betonar en viss tröskel för befolkningstätheten. Stora epidemier kan uppstå över men inte under detta tröskelvärde. Dessa arbeten används fortfarande i stor utsträckning i dagens epidemiologi.

Anderson Gray McKendrick föddes 1876 i Edinburgh som det sista av fem barn. Han studerade medicin vid universitetet i Glasgow där hans far var professor i fysiologi. År 1900 anslöt han sig till den indiska sjukvårdstjänsten. Innan han åkte till Indien följde han med Ronald Ross på ett uppdrag att bekämpa malaria i Sierra Leone. Därefter tjänstgjorde han i armén i 18 månader i Sudan. Vid sin ankomst till Indien utsågs han till läkare i ett fängelse i Bengalen där han försökte bekämpa dysenteri. År 1905 började han arbeta vid det nya Centrala institutet för medicinsk forskning i Kasauli (i norra Indien). Han arbetade med rabies men studerade också matematik. Efter att ha smittats av en tropisk sjukdom återvände han 1920 till Edinburgh och blev chef för *Royal College of Physicians Laboratory*.

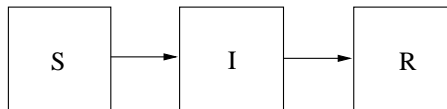
1926 publicerade McKendrick en artikel om *Matematikens tillämpning på medicinska problem*, som innehöll flera nya idéer. Han introducerade bland annat en matematisk modell för epidemier med kontinuerlig tid som tog hänsyn till den stokastiska aspekten av infektion och återhämtning.

Betrakta en population av storlek N med endast en infekterad person till en början. Människor kan successivt gå igenom tre tillstånd: det mottagliga tillståndet S , det infekterade tillståndet I och det återställda tillståndet R (figur 16.2)¹.

¹Daniel Bernoullis modell (se kapitel 4) inkluderade tillstånden S och R men inte I , eftersom infektionens varaktighet var mycket kortare än den förväntade medellivslängden.



Figur 16.1: McKendrick (1876–1943) och Kermack (1898–1970)



Figur 16.2: Möjliga tillstånd: mottaglig (S), infekterad (I), återställd (R).

Låt $p_{i,r}(t)$ vara sannolikheten för att befolkningen vid tidpunkten t innehåller exakt i personer i tillstånd I och r personer i tillstånd R , där i och r är heltal så att $1 \leq i + r \leq N$. I detta fall sägs befolkningen befinna sig i tillstånd (i, r) . Antalet mottagliga personer är $s = N - i - r$. Liksom i Ross arbete om malaria (se kapitel 12) antog McKendrick att sannolikheten för att en ny infektion skall inträffa under ett litet tidsintervall dt är lika med $asi dt$ (d.v.s. proportionell mot både antalet mottagliga personer och antalet infekterade personer). Sannolikheten för ett nytt tillfrisknande är lika med $bi dt$. Både a och b är positiva parametrar. För att beräkna $p_{i,r}(t + dt)$ bör man skilja mellan flera fall:

- populationen befinner sig i tillstånd $(i - 1, r)$ vid tiden t och en ny infektion förflyttar populationen till tillstånd (i, r) mellan t och $t + dt$; sannolikheten för denna händelse är $as(i - 1)dt$ med $s = N - (i - 1) - r$;
- populationen befinner sig i tillstånd (i, r) vid tiden t och en ny infektion förflyttar populationen till tillstånd $(i + 1, r)$ mellan t och $t + dt$; sannolikheten för denna händelse är $asi dt$ med $s = N - i - r$;

- populationen befinner sig i tillstånd $(i + 1, r - 1)$ vid tiden t och en ny återhämtning förflyttar populationen till tillstånd (i, r) mellan t och $t + dt$; sannolikheten för denna händelse är $b(i + 1)dt$;
- populationen befinner sig i tillstånd (i, r) vid tiden t och en ny återhämtning förflyttar populationen till tillstånd $(i - 1, r + 1)$ mellan t och $t + dt$; sannolikheten för denna händelse är $bi dt$.

Därför fick McKendrick fram ekvationerna

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

för $1 \leq i + r \leq N$. Den första termen på höger sida saknas när $i = 0$, medan den tredje termen saknas när $r = 0$. De initiala villkoren är $p_{i,r}(0) = 0$ för alla (i, r) utom $p_{1,0}(0) = 1$.

Med hjälp av denna modell lyckades McKendrick beräkna sannolikheten för att epidemin ska sluta med att n personer har smittats, vilket är gränsen för $p_{0,n}(t)$ när $t \rightarrow +\infty$. Det finns faktiskt inget behov av att lösa systemet (16.1). Det räcker med att notera att så länge det finns i infekterade personer och r återställda personer är sannolikheten för en ny infektion under ett litet tidsintervall dt $a(N - i - r)idt$ och sannolikheten för ett nytt återställande $bi dt$. Övergångssannolikheterna (som de vanligen kallas i teorin om Markovkedjor) från tillstånd (i, r) till tillstånd $(i + 1, r)$ eller tillstånd $(i - 1, r + 1)$ är alltså följande

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

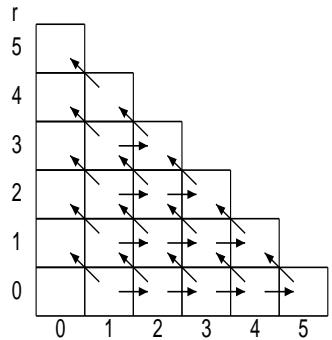
för alla $i \geq 1$ (figur 16.3).

Låt $q_{i,r}$ vara sannolikheten för att befolkningen går igenom tillstånd (i, r) under epidemin. Eftersom $i = 1$ och $r = 0$ när $t = 0$ har vi $q_{1,0} = 1$. De andra tillstånden nås antingen efter en infektion eller efter en återhämtning:

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

Den första termen på höger sida saknas när $i = 0$ eller $i = 1$. Den andra termen saknas när $r = 0$. Från denna formel kan vi först beräkna $(q_{i,0})_{2 \leq i \leq N}$, sedan $(q_{i,1})_{0 \leq i \leq N-1}$, sedan $(q_{i,2})_{0 \leq i \leq N-2}$ osv. Sannolikheten för att epidemin

Figur 16.3: Diagram som visar de möjliga tillstånden i en population med $N = 5$ (i på den horisontella axeln, r på den vertikala axeln) och de möjliga övergångarna till följd av infektion (horisontella pilar) eller återhämtning (andra pilar).



slutligen kommer att infektera n människor är $q_{0,n}$. År 1926 var sådana beräkningar ganska omständliga. Därför begränsade McKendrick sig till exempel till exempel på mycket små populationer, till exempel en familj. Med $N = 5$ personer och $b/a = 2$ fick han fram tabell 16.1. De största sannolikheterna motsvarar det fall då endast en person i familjen är smittad och det fall då hela familjen är smittad.

Tabell 16.1: Sannolikheten för att en epidemi i en familj med fem personer smittar n personer när $b/a = 2$.

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

Samma artikel från 1926 innehåller också en ny formulering av demografiska problem när tiden betraktas som en kontinuerlig variabel. Låt $P(x,t)dx$, där dx är "försvinnande liten", vara befolkningen med en ålder mellan x och $x + dx$ vid tiden t . Låt $m(x)$ vara dödligheten vid ålder x . Då

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

för ett försvinnande litet h . Introducera de partiella derivatorerna av funktionen $P(x,t)$:

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

Med hjälp av detta erhålls

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

McKendrick fick fram följande partiella differentialekvation:

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

En sådan ekvation förekommer naturligt i befolkningsproblem som struktureras av en kontinuerlig variabel, t. ex. ålder i demografi (se kapitel 25) eller tid sedan infektion i epidemiologi.

1921 utsågs William Ogilvy Kermack till chef för den kemiska avdelningen vid *Royal College of Physicians Laboratory* i Edinburgh. Kermack föddes 1898 i en liten stad i Skottland. Han studerade vid Aberdeen universitet och började forska inom organisk kemi i ett industrilaboratorium i Oxford. Trots att han blev helt blind efter en explosion i sitt laboratorium i Edinburgh 1924 fortsatte han sitt kemiska arbete med hjälp av kollegor och studenter. Kermack började också samarbeta med McKendrick om matematisk modellering av epidemier. Med början 1927 publicerade de tillsammans en serie *Bidrag till den matematiska teorin om epidemier* där de studerade deterministiska epidemimodeller. Låt N vara befolkningsstorleken med N tillräckligt stor. Anta som i 1926 års artikel att människor kan vara antingen mottagliga, smittade eller återställda. Om sjukdomen är dödlig är det tredje tillståndet i själva verket döden. Låt $S(t)$, $I(t)$ och $R(t)$ vara antalet människor i vart och ett av de tre tillstånden. Modellen är (i förenklad form) ett system av tre differentialekvationer:

$$\frac{dS}{dt} = -aSI, \quad (16.2)$$

$$\frac{dI}{dt} = aSI - bI, \quad (16.3)$$

$$\frac{dR}{dt} = bI. \quad (16.4)$$

Antalet nya infektioner per tidsenhet är därför, precis som i 1926 års stokastiska modell, proportionellt mot både antalet mottagliga personer och antalet infekterade personer. I början av epidemin, vid tiden $t = 0$, är ett visst antal personer smittade: $S(0) = N - I_0$, $I(0) = I_0$ och $R(0) = 0$, med antagandet $0 < I_0 < N$.

Även om systemet (16.2)–(16.4) inte har någon sluten lösning kan flera av dess egenskaper bevisas:

- den totala befolkningen $S(t) + I(t) + R(t)$ förblir konstant och lika med N ;
- $S(t)$, $I(t)$ och $R(t)$ förblir icke-negativa (vilket de bör vara eftersom det rör sig om populationer);
- När $t \rightarrow +\infty$ minskar $S(t)$ till gränsvärdet $S_\infty > 0$, $I(t)$ går mot 0 och $R(t)$ ökar till gränsvärdet $R_\infty < N$;
- Dessutom gäller formeln

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

som implicit ger S_∞ och därmed också den slutliga epidemins storlek, $R_\infty = N - S_\infty$.

Vi ser först att $\frac{d}{dt}(S + I + R) = 0$. Så $S(t) + I(t) + R(t) = S(0) + I(0) + R(0) = N$. Ekvationerna (16.2) och (16.3) kan skrivas om som

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

Det följer på ena sidan att $S(t) = S(0) e^{-a \int_0^t I(\tau) d\tau} > 0$ och på den andra sidan att $I(t) = I(0) e^{a \int_0^t S(\tau) d\tau - bt} > 0$. Ekvationerna (16.2) och (16.4) visar sedan att funktionen $S(t)$ är avtagande och att funktionen $R(t)$ är ökande (särskilt $R(t) \geq 0$). Eftersom $S(t) \geq 0$ och $R(t) \leq N$ har funktionerna $S(t)$ och $R(t)$ gränser när $t \rightarrow +\infty$. Eftersom $I(t) = N - S(t) - R(t)$ antar också $I(t)$ ett gränsvärde när $t \rightarrow +\infty$, som bara kan vara noll, vilket framgår av integreringen (16.4). Ekvation (16.2) visar också att

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

Genom att integrera mellan $t = 0$ och $t = +\infty$ finner vi

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

Ekvation (16.3) kan skrivas om till

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

Genom att integrera mellan $t = 0$ och $t = +\infty$ får vi

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

Genom att kombinera de två resultaten får vi formeln (16.5) som visar att $S_\infty > 0$.

När det ursprungliga antalet infekterade personer I_0 är litet jämfört med befolkningsstorleken N , vilket ofta är fallet i början av en epidemi i en stad, kan formeln (16.5) skrivas om med hjälp av $S_\infty = N - R_\infty$ som

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

där, per definition,

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

Ekvation (16.6) har en positiv lösning endast om $\mathcal{R}_0 > 1$. Kermack och McKendrick drar därför följande slutsats: Epidemin infekterar en icke försumbar del av befolkningen endast om $\mathcal{R}_0 > 1$. Det finns ett tröskelvärde för befolkningstätheten $N^* = b/a$ under vilket epidemier inte kan uppstå.

När populationsstorleken N ligger strax över detta tröskelvärde ($N = N^* + \varepsilon$) uppstår en epidemi med liten amplitud. Det följer av (16.6) att $R_\infty \approx 2\varepsilon$. Så $S_\infty \approx N^* - \varepsilon$:

Genom att använda approximationen $-\log(1-x) \approx x + \frac{x^2}{2}$ blir ekvation (16.6)

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

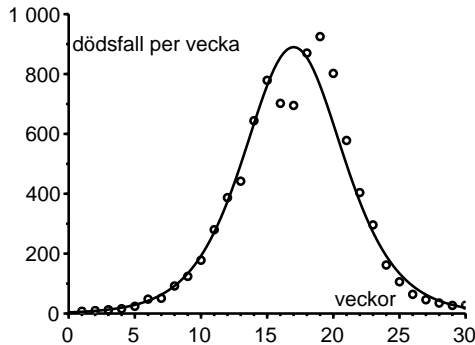
Så

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

Precis som i Ross malariamodell (kapitel 12) har villkoret $\mathcal{R}_0 > 1$ en enkel tolkning. Eftersom aN är det antal personer som en smittad person smittar per tidsenhet i början av epidemin och eftersom $1/b$ är den genomsnittliga smittoperioden, är $\mathcal{R}_0 = aN/b$ det genomsnittliga antalet sekundära fall på grund av en smittad person i början av epidemin.

För dödliga sjukdomar är $R(t)$ det kumulativa antalet dödsfall sedan epidemins början och dR/dt är antalet dödsfall per tidsenhet. Kermack och

McKendrick noterade att grafen för funktionen dR/dt i deras matematiska modell har den klockform som man förväntar sig av en epidemikurva (figur 16.4).



Figur 16.4: Kurvan dR/dt som en funktion av tiden och uppgifterna om antalet dödsfall per vecka under en pestepidemi i Bombay 1905–1906.

För att få fram dR/dt dividerade de (16.2) med (16.4) för att få

$$\frac{dS}{dR} = -\frac{a}{b} S.$$

Så

$$S(t) = S(0) \exp\left(-\frac{a}{b} R(t)\right).$$

Genom att ersätta detta i ekvation (16.4) och använda

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

fick de ekvationen

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

som fortfarande inte kan lösas explicit. Men om $\frac{a}{b} R(t)$ förblir liten under hela epidemin, ger approximationen $\exp(-u) \approx 1 - u + u^2/2$ följande resultat

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

Detta är en så kallad Riccati-ekvation med två konstanta lösningar, en positiv R_+ och en negativ R_- , som ges av rötterna till andra ordningens polynom i R på höger sida av (16.8). Låt $\tilde{R}(t)$ vara den exakta lösningen av (16.8) och sätt $Q(t) = \tilde{R}(t) - R_+$. Då uppfyller $Q(t)$ en Bernoulli-differentialekvation som liknar dem som Daniel Bernoulli och Verhulst stötte på (se (4.5) och (6.1)). Man kan alltså direkt anpassa formel (6.2) för att få $Q(t)$. En enkel men omständlig beräkning visar att dQ/dt är av formen

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

där α , β och γ är konstanter som beror på ett komplicerat sätt på modellens parametrar. Eftersom $dR/dt \approx d\tilde{R}/dt = dQ/dt$ kunde Kermack och McKendrick välja (α, β, γ) för att anpassa sina data. Naturligtvis kan moderna datorer och programvaror enkelt lösa differentialekvationen (16.7) numeriskt utan att gå igenom dessa approximationer.

Den kurva för dR/dt som man fick fram på detta sätt passade väl till uppgifterna om antalet dödsfall per vecka under pestepidemin i Bombay mellan december 1905 och juli 1906 (figur 16.4).

Kermack och McKendrick har också betraktat en mer allmän modell där smittsamheten $a(x)$ beror på tiden x sedan infektionen och där återhämningsgraden $b(x)$ också beror på x . Ekvationen som ger den slutliga epidemiska storleken (när det ursprungliga antalet smittade fall är litet) är fortfarande (16.6) men med

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

Parametern \mathcal{R}_0 har samma tolkning som i det föregående fallet: Det är det genomsnittliga antalet sekundära fall till följd av en smittad person i början av epidemin. Lägg märke till likheten mellan (16.9) och Lotsas formel (10.2) för \mathcal{R}_0 inom demografien: Ålder ersätts med tid sedan infektionen, överlevnad med sannolikheten $e^{-\int_0^x b(y) dy}$ att fortfarande vara smittad, fertilitet med kontaktfrekvensen $N a(x)$.

Kermack och McKendrick utvecklade flera andra matematiska modeller för epidemier under 1930-talet. Dessa är fortfarande byggstenar för de flesta av de mer komplexa modeller som används i dag inom epidemiologin. Parametern \mathcal{R}_0 – vars definition generaliserades av Diekmann, Heesterbeek och Metz 1990 – spelar fortfarande en central roll i modellens analys.

McKendrick gick i pension 1941 och dog 1943. Mellan 1930 och 1933 skrev Kermack några artiklar om matematisk fysik tillsammans med William

McCrea och Edmund Whittaker, båda från matematiska avdelningen vid Edinburghs universitet. Under 1930- och 1940-talen försökte Kermacks kemistgrupp syntetisera nya molekyler med antimalariakraft, men med begränsad framgång. År 1938 var Kermack tillsammans med Philip Eggleton medförfattare till en populär bok om elementär biokemi, *Det som vi är gjorda av*. Han valdes till ledamot av *Royal Society* 1944 och blev 1949 professor i biokemi vid Aberdeens universitet. Han var senare dekan för den vetenskapliga fakulteten. Han gick i pension 1968 och dog 1970.

Ytterligare läsning

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 (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

Kapitel 17

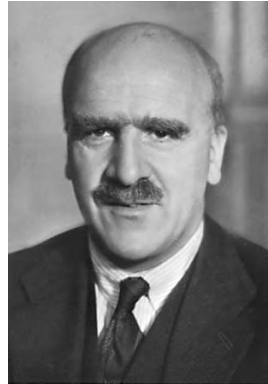
Haldane och mutationer (1927)

I en annan del av sin artikel från 1922 tog Fisher upp problemet med en mutantgen som kan överföras till ett slumpmässigt antal avkommor med en given sannolikhetsfördelning. Problemet var formellt sett detsamma som problemet med utdöende släktnamn, men i ett genetiskt sammanhang. Fisher visade att om sannolikhetsfördelningen var en Poissonfördelning och om den muterade genen inte hade någon selektiv fördel, så kunde den muterade genen försvinna ur populationen mycket långsamt. År 1927 bedrev den brittiske biologen Haldane ytterligare forskning av denna modell och visade att sannolikheten för att en muterad, fördelaktig gen skulle fortleva var dubbelt så stor som dess selektiva fördel. Han gav också en mer rigorös behandling av utrotningsproblemet.

John Burdon Sanderson Haldane föddes 1892 i Oxford, där hans far var professor i fysiologi vid universitetet. Haldane studerade vid *Eton College* och efter 1911 vid *New College* vid Oxfords universitet. Efter att ha fokuserat på matematik under sitt första år vände han sig till humaniora. Studierna avbröts av första världskriget, under vilket han tjänstgjorde i Frankrike och Irak. Efter att ha blivit sårad skickades han som militärinstruktör till Indien. År 1915 publicerade han en första artikel där han diskuterade genetiska experiment på möss som han hade påbörjat före kriget. År 1919 blev han stipendiat vid *New College*, undervisade i fysiologi och studerade, precis som sin far, andning. År 1923 anslöt han sig till F. G. Hopkins¹ biokemiska laboratorium vid Cambridge, där han fokuserade på enzymernas kinetik. Han publicerade också en science fiction-roman, *Daedalus och Icarus eller Vetenskapen och framtiden* (1923), och en essä med titeln *Callinicus, ett försvar av kemisk krigsföring* (1925). Mellan 1924 och 1934 skrev han en serie på tio artiklar med titeln *En matematisk teori om naturligt och artificiellt urval*.

I den femte artikeln i serien, som publicerades 1927, tog Haldane upp en annan genetisk modell som Fisher hade studerat 1922, en modell som fokuserade på mutationer. Fisher hade studerat sannolikheten för att en

¹Frederick Gowland Hopkins, som fick Nobelpriset i fysiologi eller medicin 1929 för sitt arbete om vitaminer.



Figur 17.1:
Haldane (1892–1964)

mutantgen ska invadera en population eller försvinna. Detta problem är formellt sett detsamma som Bienaymé, Galton och Watsons problem om utdöende släktnamn. Fisher hänvisade dock inte till dessa arbeten, även om han kan ha läst Galtons och Watsons artikel som återges i bilagan till Galtons bok *Naturlig ärfilighet* från 1889. Liksom i kapitel 9 kallar vi p_k sannolikheten för att en gen överförs till k avkommor i första generationen ($k \geq 0$). Fisher har också beaktat den genererande funktionen

$$f(x) = p_0 + p_1x + p_2x^2 + \dots + p_kx^k + \dots,$$

förutom att han inte fastställde någon övre gräns för k ; summan kan innehålla ett oändligt antal termer. Han insåg att om man utgår från en individ med den muterade genen i generation 0, är sannolikheten för att denna gen finns i k individer koefficienten för x^k i $f_1(x) = f(x)$ för generation 1, i $f_2(x) = f(f(x))$ för generation 2, i $f_3(x) = f(f(f(x)))$ för generation 3 osv. På detta sätt blir det tydligt att följande ekvation gäller.

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

Denna ekvation är mycket mer praktisk än den ekvation $f_n(x) = f_{n-1}(f(x))$ som Watson har tagit fram. I synnerhet följer det av (17.1) att utdöendesannolikheten inom n generationer $x_n = f_n(0)$ uppfyller iterationsformeln $x_n = f(x_{n-1})$, vilket Bienaymé redan hade noterat.

Som exempel tog Fisher en växt med en mutantgen som kan producera N frön, där varje frö har en sannolikhet q att överleva och producera en ny planta. Sannolikheten p_k för att få k avkommor med den muterade genen är binomial:

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

för alla $0 \leq k \leq N$ och $p_k = 0$ för $k > N$. Den genererande funktionen är då $f(x) = (1 - q + qx)^N$. Låt $\mathcal{R}_0 = Nq$ vara det genomsnittliga antalet frön som överlever för att producera en ny planta. När N är stort och q är litet, så är

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

Sannolikhetsfördelningen (p_k) går mot $e^{-\mathcal{R}_0} (\mathcal{R}_0)^k / k!$, vilket kallas en Poisson-fördelning. Fisher beräknade sedan sannolikheten för utdöende inom n generationer med hjälp av $x_0 = 0$, $x_n \approx e^{\mathcal{R}_0(x_{n-1}-1)}$ och de numeriska värdena $N = 80$ och $q = 1/80$. I detta fall är $\mathcal{R}_0 = Nq = 1$. En omständlig beräkning visar att $x_{100} \approx 0,98$: En mutantgen utan selektiv fördel ($\mathcal{R}_0 = 1$) försvinner mycket långsamt. Det finns fortfarande en chans på 2 % att genen finns kvar i populationen efter 100 generationer. År 1922 drev Fisher inte vidare studiet av denna modell.

Som en fortsättning på Fishers arbete noterade Haldane först i sin artikel från 1927 att ekvationen $x = f(x)$ för varje sannolikhetsfördelning (p_k) med $p_0 > 0$ har exakt två rötter i intervallet $(0, 1]$ när medelvärdet av antalet avkommor som bär på den muterade genen \mathcal{R}_0 är strikt större än 1, dvs. när den muterade genen har en selektiv fördel. Dessutom är utrotningssannolikheten x_∞ , som är gränsen för x_n då $n \rightarrow +\infty$, den minsta av de två rötterna till $x = f(x)$; genen har en sannolikhet som inte är noll att etablera sig i populationen. Till skillnad från Bienaymé och Cournot gav Haldane ett bevis för denna slutsats.

Faktum är att $f'(x) \geq 0$ och $f''(x) \geq 0$ på intervallet $[0, 1]$. Med andra ord är funktionen $f(x)$ växande och konvex. Antagandena $f(0) = p_0 > 0$ och $f'(1) = \mathcal{R}_0 = p_1 + 2p_2 + 3p_3 + \dots > 1$ innebär att ekvationen $f(x) = x$ har exakt två lösningar i intervallet $(0, 1]$: $x = 1$ och x^* så att $0 < x^* < 1$. Haldane hänvisade sedan till en artikel av Gabriel Koenigs från 1883, som visade att om $x_n = f(x_{n-1})$ och $x_n \rightarrow x_\infty$, så är $x_\infty = f(x_\infty)$ och $|f'(x_\infty)| \leq 1$. När $f'(1) > 1$ är den enda möjligheten att $x_\infty = x^*$.

I fallet med en Poissonfördelning med $f(x) = e^{\mathcal{R}_0(x-1)}$ och \mathcal{R}_0 som bara är något större än 1 är utdöendesannolikheten x_∞ mycket nära 1. Ekvationen $f(x_\infty) = x_\infty$ är då ekvivalent med

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

Av detta följer att

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

Haldane drog slutsatsen att sannolikheten för att den muterade genen inte ska dö ut är dubbelt så stor som dess selektiva fördel $\mathcal{R}_0 - 1$. Utan att citera Haldane tog Fisher i sin bok från 1930 som exempel ett fall där $\mathcal{R}_0 = 1,01$, vilket ger en chans på 2 % att den muterade genen inte dör ut.

Haldane blev medlem av *Royal Society* 1932. Han lämnade Cambridge för att bli professor i genetik och senare i biometri vid *University College* i London. Han var då särskilt intresserad av humangenetik: uppskattning av mutationsfrekvenser, genetiska kartor över kromosomer osv. Förutom sina vetenskapliga böcker (*Djurbiologi* 1927 tillsammans med Julian Huxley, *Enzymer* 1930 och *Orsakerna till evolutionen* 1932, *Genetikens biokemi* 1954) publicerade han ett stort antal vetenskapliga artiklar i pressen (t.ex, om livets ursprung) och några essäer (1932: *Ojämlighet mellan Människan*, 1935: *En biologs filosofi*, 1938: *Den marxistiska filosofin och vetenskaperna*, 1938: *Värdighet och politik* och 1947: *Vetenskapens framsteg*). Efter flera besök i Spanien under inbördeskriget försökte han övertyga sitt eget land att bygga skyddsrum mot flygbombningar. Under andra världskriget arbetade han med andningsproblem i ubåtar. Han var medlem av kommunistpartiet sedan 1942, men avgick 1950 på grund av det officiella förkastandet av mendelsk genetik i Sovjetunionen på grund av Lysenkos inflytande. År 1957 bosatte han sig i Indien där han fortsatte sin forskning, först vid *Indian Statistical Institute* i Calcutta och senare i Bhubaneswar. Efter att ha blivit indisk medborgare avled han 1964.

Ytterligare läsning

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)

Kapitel 18

Erlang och Steffensen om utdöendeproblemet (1929–1933)

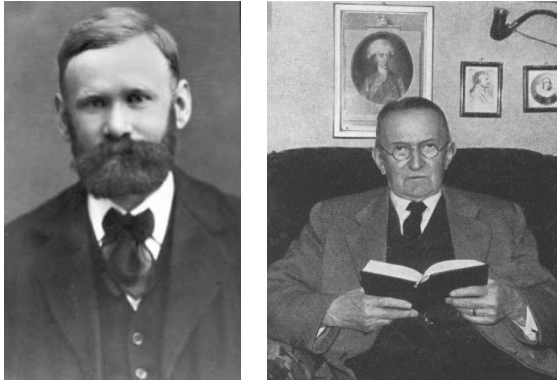
År 1929 tog den danske telefoningenjören Erlang återigen upp problemet med att släktnamn kan upphöra att existera. Hans landsman, statistikern Steffensen, utarbetade en fullständig lösning på problemet. Han visade särskilt att det förväntade antalet individer i varje generation växer exponentiellt, och skapade på så sätt en brygga mellan stokastiska och deterministiska populationsmodeller.

Agner Krarup Erlang föddes 1878 i Lønborg, Danmark. Hans far var skolmästare. Mellan 1896 och 1901 studerade den unge Erlang matematik, fysik och kemi vid Köpenhamns universitet. Därefter undervisade han under flera år i gymnasieskolor samtidigt som han behöll sitt intresse för matematik, särskilt sannolikhetsteori. Han träffade Jensen, chefsingenjör vid Köpenhamns telefonbolag och amatörmatematiker, som 1908 övertalade honom att gå med i bolagets nya forskningslaboratorium. Erlang började publicera artiklar om sannolikhetsteorins tillämpningar på hanteringen av telefonsamtal, och år 1917 upptäckte han en formel för väntetider, som snabbt användes av telefonbolag över hela världen. Hans artiklar, som först publicerades på danska, översattes sedan till flera andra språk.

1929 började Erlang intressera sig för samma problem angående utdöende som Bienaymé, Galton och Watson hade studerat före honom när det gällde släktnamn, och som Fisher och Haldane hade studerat när det gällde muterade gener. Liksom sina föregångare kände han inte till alla arbeten som hade publicerats, och genom att återigen låta p_k beteckna sannolikheten för en individ att få k barn, noterade han att sannolikheten x_n för utdöende inom n generationer är

$$x_n = p_0 + p_1 x_{n-1} + p_2 (x_{n-1})^2 + \dots = f(x_{n-1})$$

med $x_0 = 0$. Han noterade också att den totala utdöendesannolikheten x_∞ , vilken är gränsvärdet för x_n då $n \rightarrow +\infty$, är en lösning av ekvationen $x_\infty = f(x_\infty)$. Han insåg att $x = 1$ alltid var en lösning och att en annan lösning fanns mellan 0 och 1 när det genomsnittliga antalet barn $\mathcal{R}_0 = f'(1)$ är större än 1.



Figur 18.1: Erlang (1878–1929) och Steffensen (1873–1961)

Men det verkar som om han inte kunde räkna ut vilken av dessa två lösningar som var den rätta. Liksom Galton lämnade han in problemet till en tidskrift, i detta fall år 1929 till den danska *Matematisk Tidskrift*:

”Fråga 15. När sannolikheten för att en individ får k barn är p_k , där $p_0 + p_1 + p_2 + \dots = 1$, beräkna sannolikheten för att hans familj dör ut.”

Tyvärr dog Erlang samma år vid 51 års ålder. Han dog barnlös¹.

En professor i försäkringsmatematik vid Köpenhamns universitet, Johan Frederik Steffensen, tog upp Erlangs fråga. Han publicerade 1930 sin lösning i samma danska tidskrift: Sannolikheten för utdöende x_∞ är alltid den minsta roten till ekvationen $x = f(x)$ i det slutna intervallet $[0, 1]$, vilket Bienaymé och Haldane redan hade noterat. Steffensens bevis är det som återfinns i moderna läroböcker.

Vi såg att utdöendesannolikheten x_∞ är en lösning av $x = f(x)$ i det slutna intervallet $[0, 1]$. Låt x^* vara den minsta sådana lösningen. Per definition är $x^* \leq x_\infty$. Steffensen noterade först att

$$x^* = f(x^*) \geq p_0 = x_1.$$

¹Till hans minne beslutade den internationella rådgivande telefonkommittén 1946 att använda ”erlang” som namn på en enhet för att mäta intensitet i telefontrafik. Erlang är också namnet på ett programmeringsspråk från företaget Ericsson.

Anta genom induktion att $x^* \geq x_n$. Då är

$$x^* = f(x^*) \geq f(x_n) = x_{n+1}$$

eftersom funktionen $f(x)$ är ökande. Så $x^* \geq x_n$ för alla n , och det gäller även för gränsvärdet att $x^* \geq x_\infty$. Alltså är $x_\infty = x^*$.

Steffensen gav också en mer formell förklaring till varför $x = 1$ är den enda roten till $x = f(x)$ när det förväntade antalet barn $\mathcal{R}_0 = f'(1)$ är mindre än eller lika med 1 (figur 18.2a) och varför det bara finns en annan rot än $x = 1$ när $\mathcal{R}_0 > 1$ (figur 18.2b). Observera att $\mathcal{R}_0 = f'(1)$ är lutningen för funktionen $f(x)$ vid $x = 1$.

Han noterade att för varje rot i $x = f(x)$,

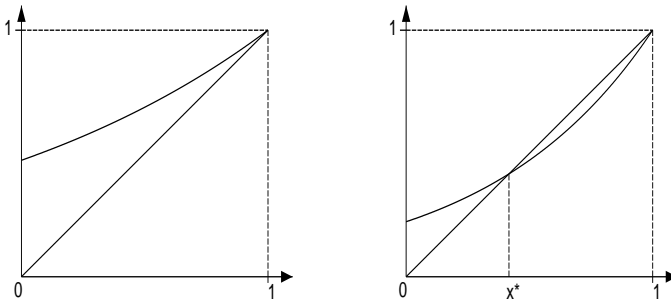
$$1 - x = 1 - f(x) = 1 - p_0 - \sum_{k=1}^{+\infty} p_k x^k = \sum_{k=1}^{+\infty} p_k (1 - x^k).$$

Om vi antar att $x \neq 1$ och dividerar med $1 - x$, får vi

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

När x ökar från 0 till 1 ökar den högra sidan av ekvation (18.1) från $1 - p_0$ till $\mathcal{R}_0 = f'(1)$. Om $\mathcal{R}_0 < 1$ har ekvation (18.1) ingen lösning. Om $\mathcal{R}_0 \geq 1$ och om vi utesluter det triviala fallet där $p_1 = 1$, så är högersidan av ekvation (18.1) en strikt växande funktion av x . Annars skulle det inte finnas något $k \geq 2$ så att $p_k \neq 0$ och \mathcal{R}_0 skulle vara lika med $p_1 < 1$. Sammanfattningsvis har (18.1) en och endast en lösning i intervallet $[0, 1]$ när $\mathcal{R}_0 \geq 1$.

Steffensen, som också var ordförande för det danska försäkrings-sällskapet och det danska matematiska sällskapet, blev inbjuden till *University of London* 1930. Hans brittiska kollega W. P. Elderton berättade för honom om Galtons och Watsons arbete. År 1933 publicerade Steffensen en ny artikel i *Annales de l'Institut Henri Poincaré*, där han hade hållit en konferens 1931. Han sammanfattade resultaten från sin artikel på danska och jämförde dem med Watsons resultat. Han visade också att väntevärdet av antalet individer i generation n är lika med $(\mathcal{R}_0)^n$.



Figur 18.2: Graf för funktionerna $y = x$ och $y = f(x)$ i exemplet i kapitel 17, $f(x) = e^{\mathcal{R}_0(x-1)}$, med $\mathcal{R}_0 = 0,75 < 1$ (vänster) eller $\mathcal{R}_0 = 1,5 > 1$ (höger).

Låt $p_{k,n}$ vara sannolikheten att det finns k individer i generation n , med utgångspunkt från en individ i generation 0. I sin artikel från 1930 hade Steffensen liksom sina föregångare noterat att den genererande funktionen

$$f_n(x) = \sum_{k=0}^{+\infty} p_{k,n} x^k$$

för generation n uppfyller $f_1(x) = f(x)$ och

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

Låt M_n vara det förväntade antalet individer i generation n . Då är

$$M_n = \sum_{k=1}^{+\infty} k p_{k,n} = f'_n(1).$$

Genom att derivera formeln (18.2) får vi

$$f'_n(x) = f'(f_{n-1}(x)) \times f'_{n-1}(x).$$

Så

$$M_n = f'_n(1) = f'(f_{n-1}(1)) \times f'_{n-1}(1) = f'(1) \times M_{n-1} = \mathcal{R}_0 \times M_{n-1}.$$

Eftersom

$$M_1 = f'_1(1) = f'(1) = \mathcal{R}_0$$

följer att $M_n = (\mathcal{R}_0)^n$ för alla n .

Det förväntade antalet individer ökar eller minskar geometriskt beroende på om \mathcal{R}_0 är större eller mindre än 1. Det förväntade antalet avkomor beter sig som i de deterministiska modellerna för befolkningstillväxt som Euler, Malthus m. fl. studerade. Även när $\mathcal{R}_0 > 1$ finns det dock en sannolikhet som inte är noll x_∞ att familjen kommer att dö ut. Denna möjlighet förekommer inte i deterministiska modeller.

Den stokastiska process som Steffensen och hans föregångare studerade är fortfarande den grundläggande beståndsdel i många mer realistiska modeller för populationsdynamik. Vi kommer att nämna detta problem en sista gång i kapitel 20. Vad gäller Steffensen förblev han professor vid Köpenhamns universitet fram till 1943, och han dog 1961.

Ytterligare läsning

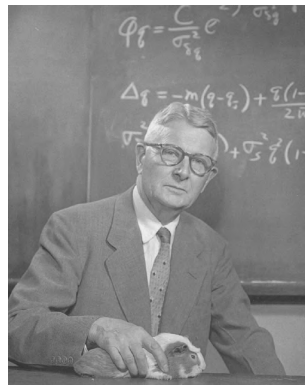
1. Brockmeyer, E., Halstrøm, H.L., Jensen, A.: The life and works of A.K. Erlang. *Trans. Dan. Acad. Techn. Sci.* 2 (1948)
2. Erlang, A.K.: Opgave Nr. 15. *Mat. Tidsskr. B*, 36 (1929) → Guttorp (1995)
3. Guttorp, P.: Three papers on the history of branching processes. *Int. Stat. Rev.* 63, 233–245 (1995) 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. Nørdlund, N.E.: Johan Frederik Steffensen in memoriam. *Nordisk Mat. Tidsskr.* 10, 105–107 (1962)
6. Ogborn, M.E.: Johan Frederik Steffensen, 1873–1961. *J. R. Stat. Soc. Ser. A* 125, 672–673 (1962)
7. Steffensen, J.F.: Om Sandssynligheden for at Afkommet uddør. *Mat. Tidsskr. B*, 19–23 (1930) → Guttorp (1995)
8. Steffensen, J.F.: Deux problèmes du calcul des probabilités. *Ann. Inst. Henri Poincaré* 3, 319–344 (1933) archive.numdam.org

Kapitel 19

Wright och slumpmässig genetisk drift (1931)

1931 utvecklade den amerikanske biologen Sewall Wright studiet av en stokastisk modell inom populationsgenetiken, som bygger på samma antaganden som Hardy-Weinberg-lagen, förutom att populationen inte antas vara oändligt stor. Genotypernas frekvenser är inte längre konstanta. En av de två allelerna kommer faktiskt att försvinna, men kanske efter mycket lång tid. Tolkningen av denna modell förblev ett föremål för tvist mellan Wright och Fisher, där den senare bedömde att det naturliga urvalet spelar en viktigare roll i evolutionen än slumpen.

Sewall Wright föddes i Massachusetts 1889. Han studerade på ett litet college i Illinois där hans far undervisade i ekonomi. Efter en masterexamen i biologi från universitetet i Illinois i Urbana och en sommarskola vid Cold Spring Harbor Laboratorium bedrev Wright forskarstudier om ärftligheten av pälsfärg hos marsvin vid Harvard. Mellan 1915 och 1925 fortsatte han att arbeta med inavelsförsök med marsvin vid djurhushållningsavdelningen vid Förenta staternas jordbruksdepartement i Washington. Han utvecklade ”metoden med stigkoefficienter” för att analysera dessa experiment. Därefter började han arbeta på zoologiska institutionen vid universitetet i Chicago.



Figur 19.1: Wright (1889–1988)

Influerad av Fishers artikel om populationsgenetik från 1922 (se kapitel 14) skrev Wright 1925 en lång artikel med titeln *Utveckling i mendelska*

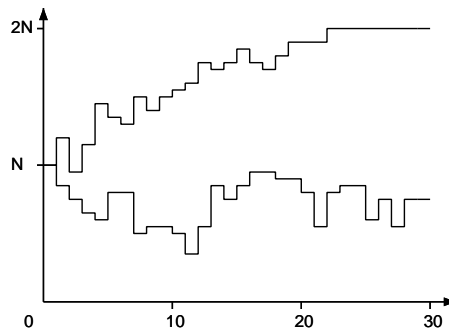
populationer, som slutligen publicerades 1931. Han studerade särskilt en matematisk modell som också förekom implicit i Fishers bok från 1930 om *Den genetiska teorin om naturligt urval*. Liksom i Hardy-Weinberg-lagen tar denna modell hänsyn till det fall där det bara finns två möjliga alleler A och a för ett locus, men populationen antas inte vara oändligt stor. Poängen är att se om borttagandet av detta antagande har något inflytande på populationens genetiska sammansättning. Låt alltså N vara det totala antalet individer, som antas vara detsamma i alla generationer. Varje individ har två alleler. Det finns alltså totalt $2N$ alleler i populationen i varje generation. I modellen antas också att parningen sker slumpmässigt. Om det finns i alleler A och $2N - i$ alleler a i generation n , kommer en allel som väljs slumpmässigt bland individer i generation $n + 1$ att vara A med sannolikhet $\frac{i}{2N}$ och a med sannolikhet $1 - \frac{i}{2N}$. Antalet alleler A i generation $n + 1$ kommer därför att vara lika med j med en sannolikhet på¹

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

där $\binom{2N}{j}$ är binomialkoefficienten som är lika med

$$\frac{(2N)!}{j!(2N-j)!}.$$

Låt X_n vara antalet alleler A i generation n ; det är en slumpmässig variabel (figur 19.2).



Figur 19.2: Två simuleringar som visar variationerna i antalet X_n alleler A under 30 generationer om $N = 20$ och $X_0 = 10$.

¹Malécot (1944) är ansvarig för denna formulering, som använder Markovkedjor.

Man kan visa att väntevärdet av X_{n+1} när man vet att $X_n = i$ är lika med i : Detta påminner om Hardy-Weinbergs lag, där frekvensen av allelen A förblev konstant genom generationer.

Betrakta den genererande funktionen

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

Väntevärdet av X_{n+1} med vetskap om att $X_n = i$ är då

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

Det är dock möjligt i denna modell att händelsen $X_n = 0$ inträffar av en slump efter ett visst antal generationer, om man utgår från ett starttillstånd $X_0 = i$ med $0 < i < 2N$. I ett sådant fall skulle alla alleler vara av typen a och X_n skulle förbli lika med 0 i alla framtida generationer. Samma fixering skulle ske med allelen A om $X_n = 2N$ efter ett visst antal generationer. Sammanfattningsvis kan man säga att när populationen antas vara oändligt stor som i Hardy-Weinberg-modellen kan de två allelerna inte försvinna eftersom deras frekvenser förblir konstanta. När man tar hänsyn till populationernas ändliga storlek, som i Fisher-Wright-modellen, fluktuerar frekvenserna för de två allelerna och en av allelerna kan (och kommer att) försvinna.

Med utgångspunkt i $X_0 = i$ kan man enkelt beräkna sannolikheten Q_i för att populationen är fixerad i tillståndet $X = 0$. Q_i måste faktiskt uppfylla "randvillkoren"

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

Dessutom är

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

eftersom $p_{i,j} Q_j$ är sannolikheten för att vara fixerad i tillståndet $X = 0$ med utgångspunkt från $X_0 = i$ och genomgång av $X_1 = j$. Eftersom

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

ser vi med hjälp av (19.2) att

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

är lösningen på systemet (19.3)–(19.4). Därför är sannolikheten för att systemet, med utgångspunkt från i alleler av typ A i en population av storlek N , utvecklas mot en population som endast innehåller allelen a lika med $1 - \frac{i}{2N}$. På samma sätt är sannolikheten att det utvecklas mot en population som endast innehåller allelen A lika med $\frac{i}{2N}$.

Wright lyckades visa att antalet generationer från start till fixering i ett av de två extrema tillstånden är i storleksordningen $2N$ generationer (figur 19.3). För populationer med flera miljoner individer skulle denna tid vara så lång att allelernas frekvenser skulle kunna betraktas som nästan konstanta, som i Hardy-Weinberg-lagen.

Anta att det finns i_0 alleler av typ A i populationen i generation 0. Låt $u_i^{(n)}$ vara sannolikheten för att det finns i alleler av typ A i populationen i generation n . Då gäller

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

för alla $j = 0, \dots, 2N$. Vi har redan sett att

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

när $n \rightarrow +\infty$ för alla $0 < i < 2N$. Wright noterade att om $u_i^{(n)} = v$ för alla $i = 1, \dots, 2N - 1$, så

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

för alla $1 < j < 2N$, eftersom $p_{0,j} = p_{2N,j} = 0$. När N är tillräckligt stort

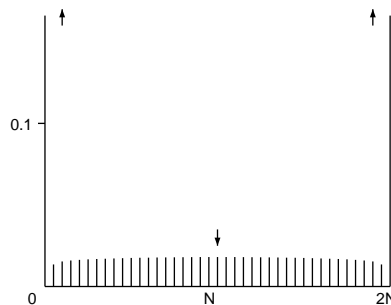
gäller

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

där värdet av integralen erhålls genom att successivt tillämpa partiell integration. Genom att kombinera (19.5) och (19.6) erhåller vi slutligen

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

för alla $0 < j < 2N$. Sannolikheterna $u_j^{(n)}$ för alla $0 < j < 2N$ minskar alltså med en hastighet av ungefär $1/2N$ per generation. Denna hastighet är mycket långsam om N är stor. Det sker nästan ingen minskning om till exempel N är i storleksordningen miljoner.



Figur 19.3: Sannolikheten att det finns i alleler A i populationen ($i = 0, \dots, 2N$ på den horisontella axeln) efter 30 generationer om $N = 20$ och $X_0 = 10$.

Fisher hade redan 1922 försökt uppskatta denna fixeringsgrad ($1/2N$) men hade missat en faktor 2. I vilket fall som helst var de två forskarna oense om den typiska storleken N på avelspopulationer. För evolutionsteorin föreslog Wrights arbete att slumpmässig genetisk drift i en liten population kunde vara en mekanism för arternas uppkomst. Biologer som arbetade med klassificering av arter hade faktiskt noterat att skillnader mellan arter eller underarter ofta inte hade någon uppenbar förklaring i form av naturligt urval. Denna idé bekämpades starkt under 1940- och 1950-talen av Fisher och hans kollega E. B. Ford som båda ansåg att slumpmässig genetisk drift

var försumbar jämfört med det naturliga urvalet. De hänvisade särskilt till sin studie av fluktuationer i genfrekvenser i en liten isolerad population av nattfjärilar (*Panaxia dominula*) nära Oxford, där de tre genotyperna för en viss gen (vanlig homozygot, heterozygot och sällsynt homozygot) kunde särskiljas genom synen. En annan berömd kontrovers om det naturliga urvalets respektive den slumpmässiga driftens inflytande gällde sniglar av släktet *Cepaea*. I mer realistiska modeller för evolutionen kombineras nu slumpmässig drift, urval, mutation, migration, icke slumpmässig parning osv. Den japanska forskaren Motoo Kimura betonade senare återigen den slumpmässiga driftens roll med sin ”neutrala teori om molekylär evolution”. En annan vidareutveckling var teorin om koalescens (introducerad av John Kingman 1982), som spårar genernas härstamning bakåt i tiden till den punkt där de har en enda gemensam förfader.

Wright blev medlem av *National Academy of Sciences* 1934. Han arbetade under många år tillsammans med Theodosius Dobzhansky med genetikerna hos naturliga populationer avflugor (*Drosophila pseudoobscura*) i Death Valley-regionen. Han gick i pension från universitetet i Chicago 1955 men fortsatte ytterligare fem år som professor vid universitetet i Wisconsin-Madison. Mellan 1968 och 1978 publicerade han en avhandling i fyra volymer som sammanfattar hans arbete om *Evolution och genetik av populationer*. Han fick Balzanpriset 1984 och dog 1988 vid 98 års ålder.

Ytterligare läsning

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. *Biog. 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)

Kapitel 20

Spridning av gener (1937)

1937 studerade Ronald Fisher och tre ryska matematiker, Kolmogorov, Petrovskij och Piskunov, oberoende av varandra en partiell differentialekvation för den geografiska spridningen av en fördelaktig gen. De visade att genfrekvensen uppförde sig som en våg som rörde sig med en väldefinierad hastighet beroende på genens selektiva fördel och en spridningskoefficient. Deras arbeten var utgångspunkten för teorin om reaktionsspridningsekvationer.

1937 publicerades två artiklar där man introducerade en ny metod för att studera rumslig variation inom populationsdynamik. Fisher var författare till den första artikeln, med titeln *Fördelaktiga geners vågutbredning*, som publicerades i *Annals of Eugenics*. Han studerade den rumsliga spridningen av en fördelaktig gen i en population. Som en förenkling betraktade han en domän som reducerades till endast en dimension och lät $u(x, t)$ beteckna den andel av populationen som befinner sig i punkten x vid tiden t och som har den gynnsamma genen. Det medför att $0 \leq u(x, t) \leq 1$. För att inkludera det naturliga urvalet använde han ekvation (14.6) med en kontinuerlig tidsvariabel,

$$\frac{\partial u}{\partial t} = au(1 - u),$$

där a är en positiv parameter. För ett givet värde på x känner vi igen detta som Verhulsts logistiska ekvation (se kapitel 6) med en lösning $u(x, t)$ som går mot 1 när $t \rightarrow +\infty$. Vidare antog Fisher att avkomman till en individ som befinner sig i punkten x med den gynnsamma genen inte stannar kvar i samma punkt utan sprids slumpmässigt i närheten av x . I analogi med fysiken hävdade han att man måste lägga till en spridningsterm till ekvationen för $u(x, t)$, vilket leder till den partiella differentialekvationen

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

När selektionskoefficienten a är noll reduceras detta till den diffusions-ekvation som introducerades av Fourier i hans värmeteorik och som senare användes av Fick för spridning av fysiska partiklar. År 1904 hade Ronald

Ross börjat beakta slumpmässig spridning inom populationsdynamiken. Han undrade då hur tätheten av myggor minskar när avståndet från en fortplantningsplats ökar. Problemet hade uppmärksammats av Karl Pearson och Lord Rayleigh. År 1937 hade den vetenskapliga litteraturen om spridningsekvationer vuxit avsevärt, särskilt efter Einsteins arbete om Brownsk rörelse.

Fisher visade att det finns lösningar till ekvationen (20.1) på formen

$$u(x, t) = U(x + vt)$$

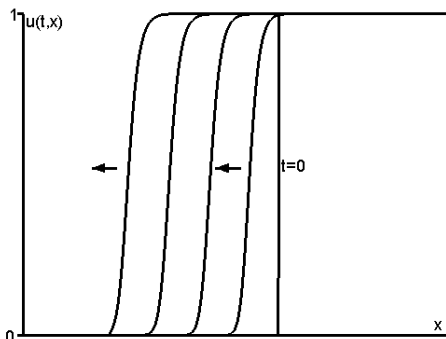
som uppfyller tre villkor:

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

förutsatt att $v \geq v^*$ där

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

Dessa lösningar förbinder det stabila tillståndet $u = 1$ där den gynnsamma genen dominerar med det stabila tillståndet $u = 0$ där genen helt saknas i populationen. Lösningarna representerar vågor som sprider sig med hastigheten v i riktning mot minskande värden på x . Faktum är att $u(x - vT, t + T) = u(x, t)$: Alltså den del av vågen som befann sig i positionen x vid tiden t flyttar sig till positionen $x - vT$ vid tiden $t + T$.



Figur 20.1: Spridning från höger till vänster av en gynnsam gen med farten v^* . Genfrekvensen $u(t, x)$ vid $t = 0$ är en stegfunktion.

Med $z = x + vt$ noterade Fisher att om $u(x, t) = U(z)$ så är $\frac{\partial u}{\partial t} = vU'(z)$, $\frac{\partial u}{\partial x} = U'(z)$ och $\frac{\partial^2 u}{\partial x^2} = U''(z)$. Om u är en lösning av ekvation (20.1), så gäller att

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

När u är nära 0, dvs. när $z \rightarrow -\infty$, förväntade sig Fisher att $U(z) \rightarrow 0$ och $U'(z) \rightarrow 0$. Om vi kallar k för gränsvärdet av $U'(z)/U(z)$ när $z \rightarrow -\infty$, vet vi från L'Hôpitals regel att $U''(z)/U'(z)$ också går mot k . Därför går

$$U''(z)/U(z) = [U''(z)/U'(z)] \times [U'(z)/U(z)]$$

mot k^2 . Genom att dividera ekvation (20.2) med $U(z)$ och låta z gå mot $-\infty$ får vi en andragradsekvation

$$Dk^2 - vk + a = 0.$$

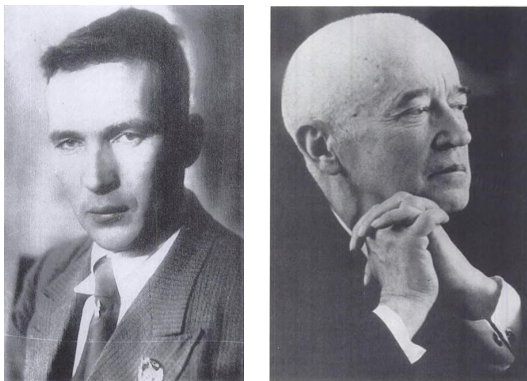
Men k måste vara ett reellt tal. Diskriminanten i denna ekvation måste alltså vara icke-negativ: $v^2 - 4aD \geq 0$, eller $v \geq 2\sqrt{aD} = v^*$. Därför är $v \geq v^*$ ett nödvändigt villkor för existensen av en våg som fortplantar sig med hastigheten v . Det är också ett tillräckligt villkor, vilket förklaras nedan.

Fisher noterade att endast den våg som utbreder sig exakt med hastigheten v^* väljs för en stor klass av startvillkor, t. ex. för stegfunktionen: $u(x, 0) = 0$ för $x < 0$, $u(x, 0) = 1$ för $x \geq 0$. Figur 20.1 visar hur detta diskontinuerliga initialtillstånd successivt blir en jämn våg som fortplantar sig i riktning mot minskande x med hastigheten v^* .

Samma år 1937, och oberoende av Fishers arbete, studerade Andrej Nikolaevich Kolmogorov, Ivan Georgievich Petrovskij och Nikolaj Semënovich Piskunov samma problem med spridning av en fördelaktig gen.

Kolmogorov föddes 1903 i Tambov i Ryssland. Under sina matematikstudier vid Moskvas statliga universitet utförde han viktiga arbeten inom trigonometriska serier. Han blev forskare vid Institutet för matematik och mekanik 1929 och universitetsprofessor 1931. Han arbetade med stokastiska processer och deras koppling till differential- och partiella differentialekvationer. År 1933 publicerade han en avhandling som lade de moderna grunderna för sannolikhetsteorin. Hans forskningsintressen omfattade topologi, approximationsteori, Markovkedjor, Brownsk rörelse och även tillämpningar på biologiska problem. År 1935 publicerade han en artikel om genetik där han diskuterade Hardys, Fishers och Wrights resultat. År 1936

publicerade han en artikel om en generalisering av Lotka-Volterra-systemet.



Figur 20.2: Kolmogorov (1903–1987) och Petrovskij (1901–1973)

Petrovskij föddes 1901 i Sevsk. Han studerade också matematik vid Moskvas statliga universitet, där han blev professor 1933. Han arbetade främst med teori för partiella differentialekvationer och topologi för reella algebraiska kurvor, men skrev också några artiklar om ordinära differentialekvationer och om sannolikhetsteori. Piskunov, som var född 1908, var en annan tidigare matematikstudent vid Moskvas statliga universitet.

Under 1930-talet hade Kolmogorov kontakt med A. S. Serebrovskij, en pionjär inom populationsgenetik i Moskva. Det blev då allt farligare att försvara den mendelska genetiken i Sovjetunionen på grund av Lysenkos dominans. Han var en agronom som hade lyckats övertyga Stalin om att den mendelska genetiken bara var ”borgerlig pseudovetenskap”. Den sjunde internationella genetikerkongressen, som ursprungligen var planerad till 1937 i Moskva, ställdes in. Många sovjetiska genetiker avrättades eller skickades till arbetsläger.

I sin artikel från 1937 med titeln *En studie av diffusionsekvationen när mängden ämne ökar och dess tillämpning på ett biologiskt problem*, som publicerades i *Moskvas statsuniversitets bulletin*, använde Kolmogorov, Petrovskij och Piskunov ändå en matematisk modell som byggde på mendelsk genetik. Deras modell var en partiell differentialekvation av formen

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

där $u(x, t)$ återigen är frekvensen av den gynnsamma genen vid punkt x och tidpunkt t . Funktionen $f(u)$ antas uppfylla ett antal villkor: $f(0) = f(1) = 0$,

$f(u) > 0$ om $0 < u < 1$, $f'(0) > 0$ och $f'(u) < f'(0)$ om $0 < u \leq 1$. Författarna visade ett resultat som är analogt med Fishers resultat men med ett mer rigoröst bevis: Om initialtillståndet är sådant att $0 \leq u(x, 0) \leq 1$, $u(x, 0) = 0$ för alla $x < x_1$ och $u(x, 0) = 1$ för alla $x > x_2 \geq x_1$, så fortplantar sig genen med hastigheten $v^* = 2\sqrt{f'(0)D}$.

Att söka efter en lösning $u(x, t) = U(z)$ där $z = x + vt$ leder till en uppenbar generalisering av ekvation (20.2), nämligen

$$vU'(z) = f(U(z)) + DU''(z).$$

Denna differentialekvation av andra ordningen kan skrivas om till ett system av differentialekvationer av första ordningen

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

Kom ihåg att $U(z)$ bör vara sådan att $U(z) \rightarrow 0$ då $z \rightarrow -\infty$, och $U(z) \rightarrow 1$ då $z \rightarrow +\infty$. I närheten av det stabila tillståndet ($U = 0, p = 0$) för systemet (20.4) har vi $f(U) \approx f'(0)U$. Det innebär att (20.4) kan approximeras med det linjära systemet

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

Om man letar efter exponentiella lösningar av formen $U(z) = U_0 e^{kz}$ och $p(z) = p_0 e^{kz}$ får man den karakteristiska ekvationen $Dk^2 - vk + f'(0) = 0$, som i Fishers artikel. Återigen måste k vara reell (annars skulle u oscillera och anta negativa värden). Således har vi att

$$v \geq 2\sqrt{f'(0)D} = v^*.$$

De två rötterna för k är då reella och positiva. Om $v > v^*$ är de två rötterna olika och det stabila tillståndet ($U = 0, p = 0$) är en instabil nod. Om $v = v^*$ är de två rötterna identiska och ($U = 0, p = 0$) är en instabil degenererad nod enligt fig. 20.3. På samma sätt leder systemet (20.4) nära det stabila tillståndet ($U = 1, p = 0$) till det linjära systemet

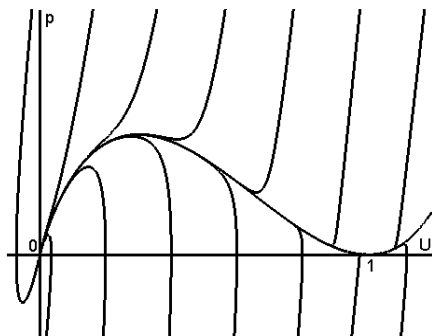
$$\frac{d(U-1)}{dz} = p, \quad \frac{dp}{dz} = \frac{vp - f'(1)(U-1)}{D}$$

och till den karakteristiska ekvationen $Dk^2 - vk + f'(1) = 0$. Diskriminanten är $v^2 - 4Df'(1) \geq 0$ eftersom $f'(1) \leq 0$. Om $f'(1) < 0$ finns det två reella rötter med motsatt tecken och $(U = 1, p = 0)$ är en sadelpunkt. Om $f'(1) = 0$ är den ena roten noll och den andra positiv (se fig. 20.3). En detaljerad analys visar att det för alla $v \geq 2\sqrt{f'(0)D}$ finns en unik integralkurva som förenar de två stabila tillstånden $(U = 0, p = 0)$ och $(U = 1, p = 0)$, som i specialfallet i fig. 20.3.

Kolmogorov, Petrovskij och Piskunov fortsatte med att visa att den partiella differentialekvationen (20.3) har en unik lösning $u(x, t)$ som uppfyller initialavillkoret, att denna lösning är sådan att $0 < u(x, t) \leq 1$ för alla x och $t > 0$, att $u(x, t)$ förblir en ökande funktion av x om den är det vid $t = 0$ och slutligen att $u(x, t)$ konvergerar mot en vågprofil som fortplantar sig med hastigheten v^* . Bevisen är för långa för att kunna sammanfattas här.

Observera att den funktion $f(u) = au(1 - u)$ som Fisher använder uppfyller alla dessa villkor med $f'(0) = a$. Inspirerade av ekvation (14.5), använde Kolmogorov, Petrovskij och Piskunov funktionen $f(u) = au(1 - u)^2$, som uppfyller samma villkor och ger samma utbredningshastighet.

Figur 20.3: Diagram (U, p) som visar några av systemets integralkurvor (20.5) och särskilt den unika kurvan som förbinder $(U = 1, p = 0)$ med $(U = 0, p = 0)$, vilken är den som ger formen på den utbredda vågen. Här gäller $f(u) = au(1 - u)^2$, $a = 1$, $D = 1$ och $v = v^* = 2$.



Artiklarna av Fisher och Kolmogorov, Petrovskij och Piskunov var utgångspunkten för konstruktionen av många matematiska modeller med geografisk spridning inom genetik, ekologi och epidemiologi. Dessa modeller är kända som "reaktions-diffusionssystem".

Kolmogorov studerade från och med 1938 även problemet med utrotning av släktnamn som Bienaymé, Galton, Watson, Fisher, Haldane, Erlang

och Steffensen tog upp: Han kallade den stokastiska process som är gemensam för alla dessa arbeten för ”förgreningsprocessen”. År 1939 blev han medlem av Sovjetunionens vetenskapsakademi. Han gjorde senare viktiga bidrag till problemet med turbulens inom flödesmekaniken (1941), till teorin om dynamiska system med koppling till celest mekanik (1953) och till informationsteori (med början 1956). Han bidrog också till att skriva ett uppslagsverk och läroböcker för gymnasier och universitet, hjälpte till att etablera en experimentell gymnasieskola och redigerade en populärvetenskaplig tidskrift. Han fick många internationella priser (bland annat Balzanpriset 1963 och Wolfpriset 1980) och avled i Moskva 1987.

Petrovskij blev dekanus för den mekaniska och matematiska fakulteten vid Moskvas statliga universitet 1940. Han var universitetets rektor från 1951 till sin död 1973. Han var fullvärdig medlem av Sovjetunionens vetenskapsakademi från 1946 och ordförande för den internationella matematikkongressen som hölls i Moskva 1966. Han skrev också läroböcker om vanliga differentialekvationer, partiella differentialekvationer och integralekvationer. Piskunov blev professor vid en militärakademi. Hans lärobok om differential- och integralkalkyl användes av många tekniska universitet. Han avled 1977.

Ytterligare läsning

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. Shiryayev, A.N.: *Selected Works of A. N. Kolmogorov*, vol. 2. Kluwer (1992)
7. Shiryayev, A.N.: Andrei Nikolaevich Kolmogorov (April 25, 1903 to October 20, 1987). In: *Kolmogorov in Perspective*, 1–88. Am. Math. Soc. (2000)

Kapitel 21

Leslie-matrisen (1945)

1945 analyserade den brittiske ekologen P. H. Leslie en matrismodell för en åldersstrukturerad population av gnagare och anpassade därmed Lotkas arbete till en diskretiserad tidsram. Han betonade att tillväxthastigheten motsvarar ett egenvärde och den stabila åldersstrukturen en egenvektor. Han uppskattade också numeriskt nettoreproduktionstakten \mathcal{R}_0 för brunråttan.

Patrick Holt Leslie föddes år 1900 nära Edinburgh i Skottland. Han studerade vid *Christ Church College* vid Oxfords universitet och fick 1921 en kandidatexamen i fysiologi. Men han kunde inte avsluta sina medicinska studier på grund av hälsoproblem. Efter att ha arbetat några år som assistent i bakteriologi vid patologiavdelningen vände han sig till statistiken och anslöt sig 1935 till *Bureau of Animal Population*, ett nytt forskningscentrum som Charles Elton hade inrättat. Syftet med centret var att studera fluktuationer i djurpopulationer genom fältstudier och laboratorieexperiment. Den största delen av forskningen gällde gnagare: analys av cyklerna hos haren och dess rovdjur lodjuret med hjälp av arkiven från *Hudson's Bay Company* i Kanada, uppföljning av den grå ekorrens territoriella expansion på bekostnad av den röda ekorren i England, insamling av uppgifter om sorkar i närheten av Oxford och så vidare. Leslie tillämpade de metoder som Lotka utvecklade för människans demografi på uppgifterna om sorkar. Under andra världskriget inriktade sig centrumets forskning på metoder för att bekämpa råttor och möss i silos.

1945 publicerade Leslie sin mest kända artikel i tidskriften *Biometrika*, som grundades av Galton, Pearson och Weldon 1901. Artikelns titel var *Om användningen av matriser i viss populationsmatematik*. Leslie undersökte en modell för tillväxten av antalet honor i en djurpopulation, t. ex. en population av råttor (men det kunde också vara människor). Populationen är uppdelad i $K + 1$ åldersgrupper: $P_{k,n}$ är antalet honor i åldern k vid tidpunkt n ($k = 0, 1, \dots, K; n = 0, 1, \dots$). Låt f_k beteckna fertiliteten vid ålder k , eller mer exakt antalet döttrar som föds per kvinna mellan tidpunkt n och tidpunkt $n + 1$. Då är K den högsta åldern med en fertilitet som inte är noll ($f_K > 0$). Låt s_k vara sannolikheten att överleva åtminstone till åldern $k + 1$ för ett djur i



Figur 21.1: P. H. Leslie (1900–1972)

åldern k . Populationens åldersstruktur ges då av följande ekvationer:

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

Alla tal f_k är icke-negativa, medan s_k uppfyller $0 < s_k < 1$. I slutet av 1800-talet och början av 1900-talet hade matematikerna tagit för vana att skriva sådana ekvationssystem i den förkortade formen¹

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

där P_n är kolumnvektorn $(P_{0,n}, \dots, P_{K,n})$ och M är den kvadratiske matrisen (dvs. taltabellen med $K + 1$ rader och $K + 1$ kolumner),

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

För att förstå hur systemet (21.1) beter sig som en funktion av tiden sökte Leslie efter en geometriskt ökande eller minskande lösning $P_n = r^n V$. Talet r och vektorn V måste uppfylla följande villkor:

$$M V = r V. \quad (21.2)$$

¹Detta innebär att $P_{k,n+1} = M_{k,0} P_{0,n} + M_{k,1} P_{1,n} + \dots + M_{k,K} P_{K,n}$ för alla k .

I detta fall kallas r för ett egenvärde och V för en egenvektor till matrisen M . Problemet är med andra ord att hitta åldersfördelningen V som vid varje tidssteg multipliceras med en konstant r . Enligt Lotkas terminologi kallas sådana fördelningar för stabila. Om vi återgår till mer vanliga beteckningar kan ekvation (21.2) skrivas om till följande:

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

Av de sista K -ekvationerna följer att

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

Genom att sätta in detta i den första ekvationen, förkorta V_0 och multiplicera med r^K får Leslie den karakteristiska ekvationen

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

Detta är en polynomekvation i r av graden $K+1$. Det finns alltså $K+1$ reella eller komplexa rötter r_1, \dots, r_{K+1} . Leslie noterade dessutom (med hjälp av Descartes teckenregel för polynom) att det bara finns en enda reell positiv rot. Vi kallar den r_1 .

Leslie föreslog också att under de flesta biologiskt realistiska förhållanden (som kan preciseras med hjälp av Perrons och Frobenius teori för icke-negativa matriser) är egenvärdet r_1 strikt större än absolutbeloppet av alla andra reella eller komplexa egenvärden (kalla dem r_2, \dots, r_{K+1}). Dessutom är alla rötter i (21.3) vanligtvis olika. För varje egenvärde r_i kan man hitta en tillhörande egenvektor. Låt Q vara en kvadratisk matris av storlek $K+1$ vars $K+1$ kolumner innehåller de egenvektorer som är associerade med r_1, \dots, r_{K+1} . Då gäller att $MQ = QD$, där D är diagonalmatrisen $[r_1, \dots, r_{K+1}]$. Så $M = QDQ^{-1}$ och

$$P_n = M^n P_0 = QD^n Q^{-1} P_0.$$

Notera att D^n är den diagonala matrisen $[(r_1)^n, \dots, (r_{K+1})^n]$ och att

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

när $n \rightarrow +\infty$ eftersom $r_1 > |r_i|$ för $i \neq 1$. Därför konvergerar $P_n / (r_1)^n$ mot $Q\mathcal{D}Q^{-1}P_0$.

Varje komponent i åldersstrukturvektorn P_n ökar eller minskar som $(r_1)^n$. Om $r_1 > 1$ ökar populationen exponentiellt. Om $r_1 < 1$ minskar den exponentiellt. Från ekvation (21.3) kan man lätt visa att villkoret $r_1 > 1$ är sant om och endast om parametern \mathcal{R}_0 , definierad av

$$\mathcal{R}_0 = f_0 + s_0 f_1 + s_0 s_1 f_2 + \cdots + s_0 s_1 \cdots s_{K-1} f_K,$$

är strikt större än 1. Observera att $s_0 s_1 \cdots s_{k-1}$ är sannolikheten att överleva till minst ålder k . Parametern \mathcal{R}_0 är alltså det genomsnittliga antalet döttrar som föds av en kvinna under hela hennes liv och är analog med formlerna (10.2), (12.2) och (16.9). Den nuvarande modellen är en slags diskret tidsanalog till Lotkas arbete (se kapitel 10) och en generalisering av Eulers arbete till åldersberoende fertiliteter (se kapitel 3).

Leslie illustrerade sin metod med hjälp av uppgifter som en amerikansk kollega publicerat om fertilitets- och överlevnadskoefficienterna f_k och s_k för brunrättan. Efter några statistiska operationer för att komplettera uppgifterna på ett rimligt sätt fick han fram $\mathcal{R}_0 \approx 26$.

Leslies matrisformulering av problem inom populationsdynamik används nu av många biologer. Beräkningarna förenklas avsevärt av moderna datorer och vetenskaplig programvara som kan beräkna egenvärden och egenvektorer för vilken matris som helst. Man kan enkelt beräkna både parametern \mathcal{R}_0 och tillväxthastigheten r_1 .

Efter andra världskriget använde Leslie sin metod för att beräkna tillväxten för andra djurarter: fåglar, skalbaggar osv. Han arbetade också med stokastiska modeller, modeller för konkurrens mellan arter och analys av fångst- och återfångstdata. Han gick i pension 1967. Samma år, då Charles Elton också gick i pension, upphörde *Bureau of Animal Population* att existera som ett oberoende forskningscentrum och blev en del av Zoologiska institutionen vid Oxfords universitet. Leslie dog 1972.

Ytterligare läsning

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)

Kapitel 22

Perkolation och epidemier (1957)

År 1957 undersökte Hammersley och Broadbent hur en ”vätska” kan spridas i ett oändligt regelbundet fyrkantigt nätverk där två angränsande noder är anslutna med en given sannolikhet. Bland de möjliga exemplen nämnde de spridningen av en epidemi i en fruktträdgård. De visade att det finns en kritisk sannolikhet under vilken ingen stor epidemi kan uppstå och över vilken stora epidemier uppstår med positiv sannolikhet. Deras artikel var utgångspunkten för perkolationsteorin.

John Michael Hammersley föddes 1920 i Skottland, där hans far arbetade för ett amerikanskt företag som exporterade stål. Han började studera vid *Emmanuel College* vid universitetet i Cambridge, men var tvungen att gå med i armén 1940. Han arbetade med att förbättra beräkningar för artilleri. Efter att ha avslutat sina studier 1948 blev han assistent vid Oxfords universitet i den grupp som arbetade med utformning och analys av experiment. År 1955 började han arbeta vid institutet för atomenergiforskning i Harwell nära Oxford.



Figur 22.1:
Hammersley (1920–2004)

Simon Ralph Broadbent föddes 1928. Han studerade ingenjörsvetenskap i Cambridge, matematik vid *Magdalen College* i Oxford (där han också skrev poesi) och påbörjade en doktorexamen i statistik vid *Imperial College* i London om *Några tester av avvikelser från enhetlig spridning*. Under sin

doktorandtid fick han stöd från den brittiska *British Coal Utilisation Research Association* för att undersöka statistiska problem som kunde relateras till kolproduktion.

1954 hölls ett symposium om Monte Carlo-metoder som sponsrades av institutet för atomenergiforskning vid *Royal Statistical Society* i London. Dessa metoder, som under 1940-talet inleddes av John von Neumann, Stanisław Ulam och Nicholas Metropolis vid Los Alamos Laboratorium, använder sig av stokastiska datorsimuleringar för att uppskatta okända matematiska storheter. Hammersley presenterade vid symposiet i London en artikel som han hade utarbetat i samarbete med Morton, en kollega från Harwell. Artikeln publicerades också i *Journal of the Royal Statistical Society*. Under diskussionen efter presentationen vid symposiet nämnde Broadbent ett intressant problem som skulle kunna studeras med hjälp av någon Monte Carlo-metod: Givet ett regelbundet nätverk av porer i två eller tre dimensioner, så att två angränsande porer är förbundna med en sannolikhet p , hur stor andel av nätverket skulle fyllas av en gas om den introducerades genom en av dessa porer? Broadbent tänkte i själva verket på utformningen av gasmasker för gruvarbetare och i synnerhet på hur stora porer som krävs för att de ska fungera.

Hammersley började sedan arbeta tillsammans med Broadbent med problemet med gasmasken. De insåg att det bara var en prototyp för en familj av problem som ännu inte hade studerats: den deterministiska spridningen av en "vätska" (betydelsen beror på sammanhanget) i ett slumpmässigt medium. Hammersley kallade det för "perkolation", i analogi med vad som händer i en kaffekanna. Vid institutet för atomenergiforskning hade Hammersley också tillgång till några av tidens mest kraftfulla datorer för att testa Monte Carlo-metoder på perkolationsproblem.

1957 publicerade Broadbent och Hammersley slutligen den första artikeln om den matematiska teorin om perkolation. Bland de exempel som de tog upp var ett av dem en modell för populationsdynamik, nämligen spridningen av en epidemi i en fruktträdgård. Träden i en mycket stor fruktträdgård antas vara placerade vid noderna i ett kvadratisk nätverk. Alla de fyra närmaste träden till ett givet smittat träd har en sannolikhet p att också vara smittade. Frågan är om ett stort antal träd kommer att smittas eller om epidemin kommer att förbli lokal. Detta beror naturligtvis på sannolikheten p , som i sin tur är kopplad till avståndet mellan träden, dvs. nätverksmaskornas bredd.

Broadbent och Hammersley undersökte det begränsande fallet där fruktträdgården är oändlig och täcker hela planet, med bara ett infekterat träd i början. Låt $f(p)$ vara sannolikheten för att ett oändligt antal träd blir infekterade från denna källa. Man förväntar sig att $f(p)$ är en ökande funktion

av p med $f(0) = 0$ och $f(1) = 1$. Deras viktigaste resultat var att det finns en kritisk sannolikhet p^* , $0 < p^* < 1$, så att

- om $p < p^*$ är $f(p) = 0$, vilket innebär att endast ett begränsat antal träd är infekterade, och
- om $p > p^*$ är $f(p) > 0$ och ett oändligt antal träd kan vara infekterade.

Beviset innebär en jämförelse med antalet olika "självundvikande promenader" i planet med utgångspunkt från smittkällan. Dessa promenader går genom ett visst antal granträd (varje träd har fyra grannar) utan att besöka något träd mer än en gång. En självundvikande promenad med n steg är en smittväg med sannolikhet p^n eftersom infektionen kan överföras från varje besökt träd till nästa med sannolikhet p . Låt nu $q(j, n)$ vara sannolikheten för att det bland alla självundvikande promenader med n steg finns exakt j sådana promenader som är smittvägar. Om det finns ett oändligt antal infekterade träd, finns det för alla heltal n minst en självundvikande promenad med n steg som är en infektionsväg. Så

$$0 \leq f(p) \leq \sum_{j=1}^{+\infty} q(j, n) \leq \sum_{j=1}^{+\infty} j q(j, n)$$

för alla n . Men

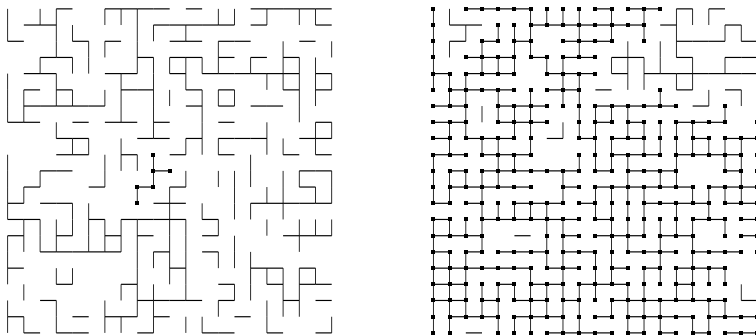
$$\sum_{j=1}^{+\infty} j q(j, n)$$

är det förväntade antalet självundvikande promenader med n steg som är smittvägar. Detta antal är lika med $p^n s(n)$, där $s(n)$ är det totala antalet självundvikande promenader med n steg. Hammersley kunde i en kompletterande artikel visa att $s(n)$ växer som $e^{\kappa n}$ som $n \rightarrow +\infty$, där κ kallas konnektivitetskonstanten. Om $p < e^{-\kappa}$ så går $p^n s(n)$ mot 0 då $n \rightarrow +\infty$, och $f(p) = 0$. Således är $p^* \geq e^{-\kappa} > 0$.

I praktiken är det därför bättre om träden inte står för nära varandra för att hålla p under p^* i händelse av en epidemi. Men ju närmare träden står varandra, desto högre produktion per hektar. En kompromiss måste hittas.

Som Broadbent och Hammersley noterade finns det en viss likhet mellan existensen av en kritisk sannolikhet i perkolationsprocesser och existensen av en tröskel i förgreningsprocesser (se kapitel 7).

Man kan försöka uppskatta den kritiska sannolikheten p^* numeriskt. För detta ändamål fastställer man ett värde för p och approximerar det oändliga nätverket med ett ändligt kvadratisk nätverk av storleken $N \times N$ med N tillräckligt stort. Anta till exempel att trädet i nätverkets mitt är infekterat. Med en dator kan man slumpmässigt välja vilka träd som kan infektera andra träd. Figur 22.2 visar de slumpmässigt valda infektionsvägarna med hjälp av kanter som i en graf. I figur 22.2a är p mindre än p^* . I figur 22.2b är p större än p^* . Man kan lätt bestämma vilka träd som kan infekteras, nämligen de som kan nås genom en väg av kanter som börjar från det infekterade trädet i mitten. De är markerade med små svarta fyrkanter i figurerna.

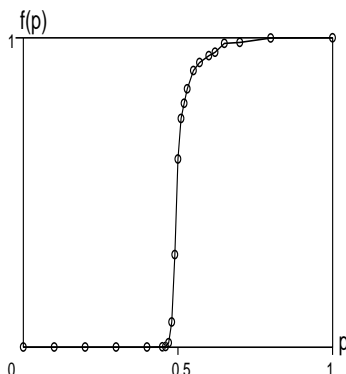


Figur 22.2: (a) Perkolation med $p = 0,4$. (b) Perkolation med $p = 0,55$.

Man kan sedan kontrollera om epidemin har nått åtminstone gränsen för $N \times N$ -nätverket. Om så är fallet och om N är tillräckligt stort kan man anse att antalet infekterade träd är ”nästan oändligt”. Genom att upprepa denna typ av simulering många gånger kan man hitta ett ungefärligt värde på sannolikheten $f(p)$ att antalet infekterade träd är oändligt (detta är Monte Carlo-metoden). Genom att låta p variera mellan 0 och 1 kan man slutligen få en approximation av tröskelvärdet p^* , som är det minsta värdet så att $f(p) > 0$ om $p > p^*$.

Broadbent och Hammersleys artikel innehöll endast ett bevis för att tröskelvärdet p^* existerar. Under de följande åren fortsatte Hammersley att utveckla den matematiska teorin om perkolation, medan Broadbent vände sig till andra ämnen. I och med datorernas utveckling på 1970-talet blev det lättare att utföra de simuleringar som beskrivs ovan (figur 22.3). Man antog

då att $p^* = 1/2$. Detta resultat bevisades slutligen 1980 av Harry Kesten från Cornell universitet.



Figur 22.3: Sannolikheten $f(p)$ att oändligt många träd blir infekterade som en funktion av p . Kurvan erhålls genom att köra 1 000 simuleringar på ett nätverk med 200×200 .

Mellan 1959 och 1969 arbetade Hammersley för Institutet för ekonomi och statistik vid Oxfords universitet. Han blev ledamot av *Trinity College*. År 1964 publicerade han i samarbete med David Handscomb en bok med titeln *Monte Carlo-metoder*. Han valdes in i *Royal Society* 1976. Han gick i pension 1987 men fortsatte att besöka centret för industriell och tillämpad matematik i Oxford. Han dog 2004.

Broadbent disputerade vid *Imperial College* 1957. Han fick jobb på ett industriföretag, *United Glass Bottle Manufacturers*. Efter tio år inom industrin började han arbeta på en nyhetsbyrå, *London Press Exchange*, som gjorde vetenskapliga läsundersökningar. Byrån köptes 1969 av *Leo Burnett*, ett amerikanskt reklamföretag. Broadbent arbetade med hur man kan mäta reklamens effektivitet och publicerade flera böcker i ämnet: *Att spendera reklampengar* (1975), *Budget för reklam* (1989), *Redovisningsbar reklam* (1997) och *När ska du annonsera?* (1999). 1980 var han med och grundade *Advertising Effectiveness Awards*. Han tillbringade flera som direktör år på Leo Burnetts huvudkontor i Chicago. Han drev också sin egen konsultfirma, *BrandCon Limited*. Han dog 2002.

Ytterligare läsning

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)

Kapitel 23

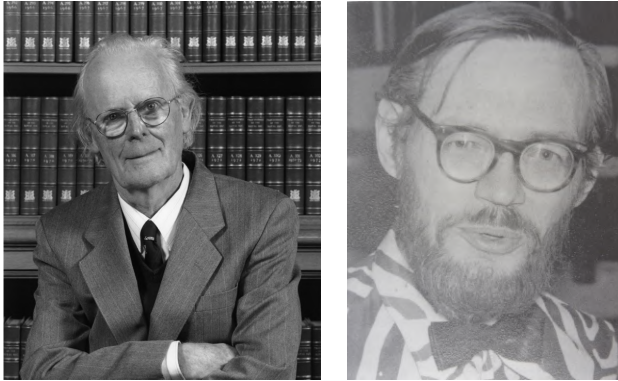
Spelteori och evolution (1973)

1973 publicerade Maynard Smith och Price en artikel där de analyserade varför djur undviker att använda sina farligaste vapen i konflikter inom arten. Deras modell använde sig av spelteori och inledde tillämpningen av denna matematiska teori på evolutionära problem.

John Maynard Smith föddes i London 1920. Hans far, som var kirurg, dog när han var åtta år. Maynard Smith studerade vid *Eton College* och började studera teknik vid *Trinity College*, Cambridge. Han var då medlem i Storbritanniens kommunistparti. När kriget bröt ut 1939 försökte han anmäla sig frivilligt till armén men fick avslag på grund av sin dåliga syn. Han fullföljde sina ingenjörstudier och arbetade under några år med konstruktion av militärflygplan. Slutligen bestämde han sig för att vända sig till biologin och studerade genetik vid *University College* i London med Haldane som handledare. Han blev docent i zoologi 1952. Han lämnade kommunistpartiet efter händelserna i Ungern 1956. Hans första bok, med titeln *Evolutionsteorin*, publicerades 1958. År 1965 blev han professor i biologi vid det nygrundade universitetet i Sussex. Han publicerade då ytterligare två böcker: *Matematiska idéer inom biologin* (1968) och *Om evolution* (1972).

George R. Price föddes 1922 i USA. Han studerade kemi vid universitetet i Chicago och disputerade 1946 efter att ha arbetat i Manhattanprojektet. År 1950 blev han biträdande forskare i medicin vid universitetet i Minnesota. Han arbetade senare som oberoende journalist för olika tidskrifter innan han återvände till forskningen på IBM. Efter att ha behandlats för sköldkörtelcancer bosatte han sig 1967 i England och började studera ett helt annat ämne: evolutionsbiologi. Han arbetade i London vid Galtonlaboratoriet vid *University College* från 1968. Hans första artikel på detta nya område, *Urval och kovarians*, publicerades med hjälp av W. D. Hamilton i ett nummer av *Nature* 1970 och innehöll det som nu kallas Prices ekvation.

Price lämnade också in ytterligare en artikel till *Nature*, den här gången om konflikter mellan djur. Men den hade inte rätt format för denna tidskrift. Maynard Smith, som var granskare, föreslog därför att man skulle göra en kortare version. Price började arbeta med något annat medan Maynard Smith



Figur 23.1: Maynard Smith (1920–2004) och Price (1922–1975)

började utveckla Prices idé på egen hand. Till slut publicerade Maynard Smith och Price en gemensam artikel med titeln *Logiken i djurkonflikter*, som *Nature* publicerade 1973. Artikeln utgjorde ett intressant bidrag till användningen av spelteori inom evolutionsbiologin. Dessförinnan hade spelteorin främst utvecklats för ekonomi och politik, särskilt efter John von Neumanns och Oskar Morgensterns bok från 1944 med titeln *Teori om spel och ekonomiskt beteende*. Maynard Smith och Price utgick från följande fråga: Hur kommer det sig att de ”vapen” som djuren har till sitt förfogande (horn, klor, gift osv.) i konflikter mellan djur av samma art sällan används för att döda? Enligt Darwins idéer om kampen för livet borde mer aggressiva djur vinna fler strider och få ett större antal avkommor, vilket leder till en eskalering av användningen av ”vapen”. Lägg märke till att detta var under det kalla kriget, så ämnet hade då också en politisk prägel.

Maynard Smith och Price föreställde sig en serie spel där två djur kan konkurrera om en resurs, till exempel ett revir i en gynnsam livsmiljö. I den förenklade presentation som Maynard Smith skulle använda i sin bok från 1982 *Evolution och spelteori* antar varje djur antingen ”hökstrategin” eller ”duvstrategin”. I det följande talar vi helt enkelt om hökar och duvor, men vi menar strategier som antas av djur av samma art. Låt $V > 0$ vara resursens värde, vilket innebär att om \mathcal{R}_0 är det normala genomsnittliga antalet avkommor för ett djur, har vinnaren av tävlingen i genomsnitt $\mathcal{R}_0 + V$ avkommor.

Om en hök möter en annan hök slåss de om resursen: Vinnaren får resursen till ett värde av V , förloraren drabbas av en ”kostnad” av $C > 0$. Var och en av de två hökarna har en sannolikhet lika med $1/2$ att vinna

tävlingen och samma sannolikhet att förlora. Den förväntade vinsten från en kamp mellan två hökar är därför $\frac{1}{2}(V - C)$ för de två konkurrenterna. Om en hök däremot möter en duva, får hökaren resursen V medan duvan flyr utan att slåss och kostnaden är 0. Om två duvor möts får den ena av dem resursen V , den andra flyr utan att slåss och utan kostnad. Eftersom båda duvorna har samma sannolikhet $1/2$ att vinna, är den förväntade vinsten när två duvor möts därför $V/2$. Utbetalningarna kan sammanfattas som i tabell 23.1.

Tabell 23.1: Förväntade vinster i spelet hök-duva.

	en hök	en duva
vinster av en hök mot ...	$\frac{1}{2}(V - C)$	V
vinster av en duva mot ...	0	$V/2$

Mer allmänt kan man tänka sig strider mellan individer som kan anta en av två strategier, kalla dem 1 och 2, med en matris av förväntade vinster $(G_{i,j})_{1 \leq i,j \leq 2}$. I exemplet ovan följer hökar strategi 1, duvor strategi 2, $G_{1,1} = \frac{1}{2}(V - C)$, $G_{1,2} = V$, $G_{2,1} = 0$ och $G_{2,2} = V/2$. I den ursprungliga artikeln från 1973 hade Maynard Smith och Price faktiskt redan använt datorsimuleringar för att testa mer än två möjliga strategier (dessa kallades "hök", "mus", "mobbare", "hämnare" och "test-hämnare").

Föreställ dig nu en stor population av djur av samma art med en andel x_n hökar och en andel $1 - x_n$ duvor i generation n . Hökarna i generation n har ett genomsnittligt antal avkomor som är lika med

$$R_1(n) = \mathcal{R}_0 + x_n G_{1,1} + (1 - x_n) G_{1,2}. \quad (23.1)$$

På samma sätt har duvor ett genomsnittligt antal avkomor som är lika med

$$R_2(n) = \mathcal{R}_0 + x_n G_{2,1} + (1 - x_n) G_{2,2}. \quad (23.2)$$

Det genomsnittliga antalet avkomor i hela populationen är därför

$$R(n) = x_n R_1(n) + (1 - x_n) R_2(n).$$

Om vi bortser från eventuella subtiliteter som beror på sexuell reproduktion kan vi konstatera att andelen hökar i nästa generation är

$$x_{n+1} = x_n R_1(n) / R(n). \quad (23.3)$$

Därför är $x_{n+1} > x_n$ om $R_1(n) > R(n)$ och $x_{n+1} < x_n$ om $R_1(n) < R(n)$. Det finns tre möjliga jämviktstillstånd: $x = 0$, $x = 1$ och

$$x^* = \frac{G_{1,2} - G_{2,2}}{G_{2,1} - G_{1,1} + G_{1,2} - G_{2,2}}$$

förutsatt att $0 < x^* < 1$. I hök-duva-spelet är

$$x^* = V/C < 1$$

förutsatt att $V < C$.

Faktum är att $x = 0$ är ett uppenbart jämviktstillstånd för (23.3). Om $x \neq 0$ är ett annat jämviktstillstånd, så är $R_1 = R = xR_1 + (1-x)R_2$. Så antingen $x = 1$ eller $R_1 = R_2$. Den senare möjligheten är likvärdig med

$$xG_{1,1} + (1-x)G_{1,2} = xG_{2,1} + (1-x)G_{2,2},$$

vilket ger det jämviktstillståndet x^* .

Jämviktstillståndet $x = 1$ motsvarar en population där 100% av individerna följer strategi 1. Detta tillstånd är stabilt om det inte kan invaderas av några få individer som följer strategi 2. Av (23.3) framgår att detta villkor är likvärdigt med att $R_1(n) > R(n)$ för alla x_n tillräckligt nära 1. Eftersom $R(n) = x_n R_1(n) + (1-x_n)R_2(n)$, blir villkoret $R_1(n) > R_2(n)$ för alla x_n tillräckligt nära 1. Om vi tittar på uttrycken (23.1)–(23.2) för R_1 och R_2 kommer vi fram till slutsatsen att $x = 1$ är stabilt om och endast om ett av följande två villkor är uppfyllt:

- $G_{1,1} > G_{2,1}$;
- $G_{1,1} = G_{2,1}$ och $G_{1,2} > G_{2,2}$.

Om så är fallet sägs strategi 1 vara en *evolutionärt stabil strategi*¹. I spelet hök-duva är villkoret $G_{1,2} > G_{2,2}$ alltid sant. Så hökstrategin är evolutionärt stabil om och endast om $G_{1,1} \geq G_{2,1}$, dvs. $V \geq C$.

Jämviktstillståndet $x = 0$ motsvarar en population där alla individer följer strategi 2. Denna situation är symmetrisk till den föregående om vi byter ut index 1 och 2. I hök- och duvspelet har vi $G_{1,2} = V > G_{2,2} = V/2$ så jämviktstillståndet $x = 0$ är alltid instabilt. Att införa ett litet antal hökar i en population av duvor skulle leda till en progressiv invasion av hökarna.

På samma sätt kan man visa att det tredje jämviktstillståndet x^* , förutsatt att $0 < x^* < 1$, alltid är stabilt. I spelet hök-duva motsvarar $x^* = V/C$ en blandad population med både hökar och duvor.

Sammanfattningsvis finns det två fall i spelet hök-duva. Om $V \geq C$, dvs. om värdet av resursen är större än den möjliga kostnaden, går populationen

¹Mer exakt är jämvikten $x = 1$ "evolutionärt stabil". Skillnaden mellan jämvikt och strategi är viktig om det finns mer än två strategier.

mot ett jämviktstillstånd med hökar men inga duvor, oavsett utgångsläget $x(0)$ med $0 < x(0) < 1$. Hökstrategin är då en evolutionärt stabil strategi. Om däremot $V < C$ går populationen mot ett blandat jämviktstillstånd med en andel x^* hökar och en andel $1 - x^*$ duvor. Modellen ger alltså en förklaring till varför individer med mindre aggressiva beteenden kan överleva när $V < C$. Formeln $x^* = V/C$ visar dessutom att ju högre kostnaden C för förlorare är, desto mindre blir andelen x^* hökar i populationen. Därför använder arter med de farligaste ”vapnen” sällan dessa för strider inom arten; de föredrar oskadliga rituella strider, där konkurrerande djur försöker imponera på varandra, men undviker verkliga strider som skulle kunna orsaka skador.

Den ursprungliga artikeln från 1973 av Maynard Smith och Price diskuterade begreppet evolutionärt stabil strategi och använde huvudsakligen datorsimuleringar av djurkonflikter där man registrerade vinsterna av olika strategier. Metoden med dynamiska ekvationer som (23.3) utvecklades något senare, särskilt av Taylor och Jonker. Sedan dess har många författare tillämpat idéer från spelteorin på frågor inom evolutionsbiologin eller omvänt tillämpat dynamiska evolutionära metoder på mer klassiska problem inom spelteorin. Förutom frågor som rör konflikter mellan djur kan man till exempel nämna problem om föräldrars investeringar eller om könskvoter (förhållandet mellan antalet hanar och honor vid födseln), som redan studerades av Carl Düsing 1884 och av Ronald Fisher i hans bok från 1930 om *Den genetiska teorin om naturligt urval*. Vissa andra modeller fokuserar på de dynamiska aspekterna av ”fångarnas dilemma” eller ”sten, sax, påse”-spelet. Man insåg också att begreppet evolutionärt stabil strategi är nära besläktat med begreppet Nash-jämvikt inom spelteorin.

Price, som hade varit övertygad ateist, fick en mystisk upplevelse 1970 och konverterade till den kristna tron. Han gav upp sin forskning 1974 eftersom han ansåg att ”den typ av teoretisk-matematisk genetik som han sysslat med inte var särskilt relevant för mänskliga problem”. Han gav alla sina ägodelar till hemlösa och begick självmord några månader senare.

Maynard Smith däremot fortsatte denna linje och valdes in i *Royal Society* 1977. Han publicerade många böcker: *Modeller inom ekologi* (1974), *Utvecklingen av sex* (1978), *Evolution och spelteori* (1982), *Biologiska problem* (1986), *Hade Darwin rätt?* (1988) och *Evolutionär genetik* (1989). Han publicerade också i samarbete med E. Szathmáry *De stora övergångarna i evolutionen* (1995) och *Livets ursprung: Från livets födelse till språkets ursprung* (1999). Han gick i pension 1985. År 1999 fick han Crafoordpriset i biovetenskap av Kungliga Vetenskapsakademien för sina ”fundamentala bidrag till den konceptuella utvecklingen av evolutionsbiologin”. År 2003 publicerade han i samarbete med D. Harper *Djursignaler*. Han dog i Sussex

2004.

Ytterligare läsning

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. Harman, O.: *The Price of Altruism*. W. W. Norton, London (2010)
5. Maynard Smith, J., Price, G.R.: The logic of animal conflict. *Nature* 246, 15–18 (1973)
6. Maynard Smith, J.: *Evolution and the Theory of Games*. Cambridge University Press (1982)
7. 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
8. Sigmund, K.: John Maynard Smith and evolutionary game theory. *Theor. Pop. Biol.* 68, 7–10 (2005)
9. Taylor, P.D., Jonker, L.B.: Evolutionary stable strategies and game dynamics. *Math. Biosci.* 40, 145–156 (1978)
10. Von Neumann, J., Morgenstern, O.: *Theory of Games and Economic Behavior*. Princeton University Press (1944) archive.org

Kapitel 24

Kaotiska populationer (1974)

1974 studerade Robert May, en australiensisk fysiker som blivit ekolog, den diskreta logistiska ekvationen som en modell för populationsdynamik. Han noterade att oväntade bifurkationer uppstod och att det asymptotiska beteendet till och med kunde vara kaotiskt. Dessa observationer innebär att långsiktiga förutsägelser kan vara omöjliga även med en enkel deterministisk modell. Mays artikel var en av de artiklar som lanserade "kaosteorin".

Robert McCredie May föddes 1936 i Australien. Efter att ha studerat teoretisk fysik och tagit en doktorsexamen vid universitetet i Sydney 1959 tillbringade han två år på avdelningen för tillämpad matematik vid Harvard universitet. När han återvände till Australien blev han professor i teoretisk fysik. Då han 1971 besökte *Institute for Advanced Study* i Princeton bytte han forskningsämne och började fokusera på populationsdynamik. År 1973 blev han professor i zoologi vid Princeton. Samma år publicerade han en bok med titeln *Stabilitet och komplexitet i modell-ekosystem*.



Figur 24.1:
Robert M. May (1936–2020)

1974 publicerade May i *Science* en artikel med titeln *Biologiska populationer med icke överlappande generationer: stabila punkter, stabila cykler och kaos*, där han visade att mycket enkla matematiska modeller för populationsdynamik kan bete sig kaotiskt.

För att förstå ursprunget till detta problem måste man gå ungefär tio år tillbaka i tiden. År 1963 upptäckte Edward Lorenz, en amerikansk meteorolog som arbetade vid teknikinstitutet i Massachusetts (M.I.T.), när

han gjorde numeriska simuleringar på sin dator, att en förenklad modell av atmosfären med endast tre differentialekvationer kunde uppträda på ett mycket överraskande sätt: En liten förändring av de initiala förhållandena kunde helt förändra slutresultatet av simuleringen och därmed även de meteorologiska prognoserna. Matematikern Henri Poincaré hade efter att ha arbetat med planeternas rörelse i solsystemet faktiskt redan i början av 1900-talet, långt före datoråldern, tänkt på denna möjlighet. Men i början av 1970-talet var det bara några få forskare som undersökte denna märkliga egenskap. Vid universitetet i Maryland studerade James Yorke Lorenz arbete och introducerade begreppet "kaos" i detta sammanhang. Artikeln han skrev tillsammans med sin elev Tien-Yien Li, med titeln *Period tre innebär kaos*, publicerades 1975¹.

May å sin sida fokuserade på modellen

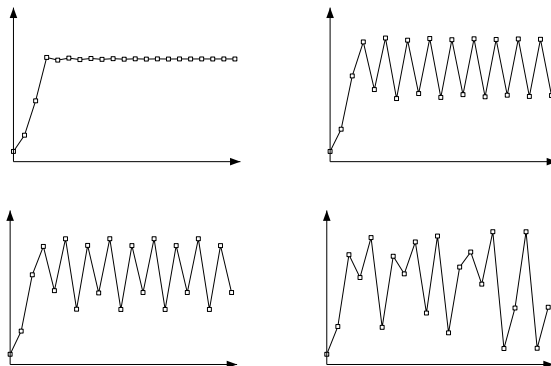
$$p_{n+1} = p_n + a p_n(1 - p_n/K), \quad (24.1)$$

där a och K är positiva parametrar och p_n står för storleken på en djurpopulation år n . När p_n är liten jämfört med bärandekapaciteten K ligger dynamiken nära en geometrisk tillväxt $p_{n+1} \approx (1 + a) p_n$. Den fullständiga ekvationen är en slags diskret tidsanalog till den logistiska ekvationen som infördes av Verhulst (se kapitel 6). Men till skillnad från den senare visade May att den diskreta tidekvationen kan ha ett mycket mer överraskande beteende, vilket är lätt att observera med en enkel fickräknare som kan utföra addition och multiplikation (figur 24.2). Maynard Smith hade redan beaktat ekvationen (24.1) i sin bok *Matematiska idéer inom biologin* från 1968. Men trots att han hade provat ett par numeriska värden för a hade han inte insett att ekvationen var speciell.

Figur 24.2, som liknar den i Mays artikel från 1974, visar att populationen p_n konvergerar mot ett stabilt tillstånd när $0 < a < 2$. När $2 < a \leq 2,449$ (den övre gränsen 2,449 är en approximation) går populationen p_n mot en cykel med period 2. När $2,450 \leq a \leq 2,544$ går populationen p_n mot en cykel med period 4. När $2,545 \leq a \leq 2,564$ går p_n mot en cykel med period 8, osv. Intervallen för parametern a för vilka p_n går mot en cykel med perioden 2^n blir mindre när n ökar och överstiger aldrig 2,570. När $a \geq 2,570$ kan p_n uppträda på ett "kaotiskt" sätt.

1976 skrev May en översikt över problemet, som publicerades i *Nature*, med titeln *Enkla matematiska modeller med mycket komplicerad dynamik*. Där samlade han inte bara sina egna resultat utan även andra forskares. Om

¹Ett mer generellt resultat bevisades av O. M. Sharkovskij 1964, men hans artikel som publicerades i en ukrainsk matematisk tidskrift var inte allmänt känd.



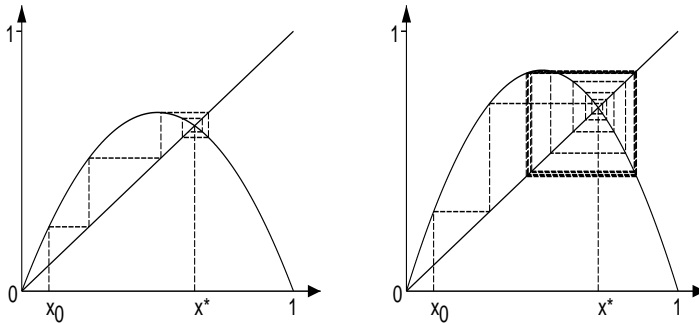
Figur 24.2: I alla figurer är n på den horisontella axeln, p_n på den vertikala axeln och $p_0 = K/10$. Linjerna erhålls genom att sammanfoga punkterna med koordinaterna (n, p_n) . Överst till vänster: $0 < a < 2$ (stationärt tillstånd). Överst till höger: $2 < a \leq 2,449$ (cykel med period 2). Nederst till vänster: $2,450 \leq a \leq 2,544$ (cykel med period 4). Nederst till höger: $2,570 \leq a \leq 3$ (eventuellt kaos).

man först låter $x_n = \frac{ap_n}{K(1+a)}$ och $r = 1 + a$ (så att $r > 1$) ser man att ekvation (24.1) kan skrivas om i den enklare formen

$$x_{n+1} = rx_n(1 - x_n). \quad (24.2)$$

För att denna ekvation ska beskriva dynamiken hos en population måste x_n vara icke-negativ för alla n . Vi antar alltså att det initiala tillståndet x_0 uppfyller $0 \leq x_0 \leq 1$ och att $r \leq 4$. Det sistnämnda villkoret säkerställer att den högra sidan av (24.2) ligger mellan 0 och 1. Det värt att notera att det kaotiska fallet $r = 4$ hade använts som en slumpalsgenerator av Stanisław Ulam och John von Neumann redan 1947. Om vi inför funktionen $f(x) = rx(1 - x)$, kan ekvation (24.2) skrivas om till $x_{n+1} = f(x_n)$ och jämviktstillstånden är lösningarna av $x = f(x)$. Grafiskt sett är detta skärningspunkterna mellan kurvorna $y = f(x)$ och $y = x$ (figur 24.3). Observera att $x = 0$ alltid är ett jämviktstillstånd. Eftersom $r > 1$ finns det också ett annat jämviktstillstånd $x^* > 0$ sådant att $x^* = rx^*(1 - x^*)$, dvs. $x^* = 1 - 1/r$.

Eftersom $r > 1$ är jämviktstillståndet $x = 0$ instabilt. När x_n är nära 0 har vi faktiskt $x_{n+1} \approx rx_n$. Så x_n tenderar att avlägsna sig från 0. Jämviktstillståndet x^* är lokalt stabilt endast för $1 < r < 3$.



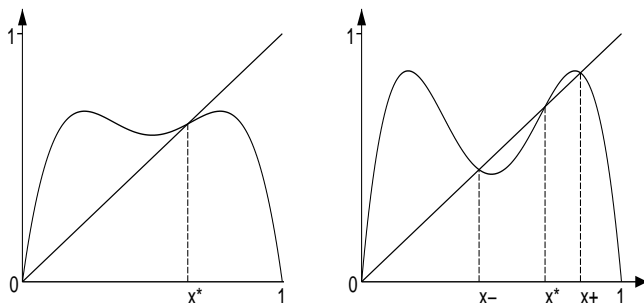
Figur 24.3: Funktionen $y = f(x) = rx(1-x)$, den raka linjen $y = x$, det stationära tillståndet x^* och sekvensen definierad av $x_{n+1} = f(x_n)$. (a) $r = 2,75$, sekvensen går mot x^* . (b) $r = 3,4$, det stabila tillståndet x^* är instabilt och sekvensen konvergerar till en cykel med period 2.

Sätt $y_n = x_n - x^*$. Då är (24.2) ekvivalent med $y_{n+1} = (2 - r - ry_n)y_n$. Om x_n är nära x^* , är y_n nära 0 och $y_{n+1} \approx (2 - r)y_n$. Men om $y_{n+1} = ky_n$, då $y_n = k^n y_0$ så att $y_n \rightarrow 0$ när $n \rightarrow \infty$ om och endast om $-1 < k < 1$. Här är det stabila tillståndet x^* lokalt stabilt om och endast om $-1 < 2 - r < 1$, dvs. $1 < r < 3$.

När $1 < r < 3$ kan man visa att för alla initialförhållanden $0 < x_0 < 1$ konvergerar talföljden x_n verkligen mot x^* (figur 24.3a). Men vad händer då $3 < r \leq 4$? För att besvara denna fråga bör man först notera att $x_{n+2} = f(x_{n+1}) = f(f(x_n))$. Introducera funktionen $f_2(x) = f(f(x)) = r^2 x(1-x)(1-rx(1-x))$ och betrakta lösningarna till ekvationen $x = f_2(x)$, som kallas fixpunkter för funktionen $f_2(x)$. Grafiskt sett är det skärningspunkterna mellan kurvorna $y = f_2(x)$ och $y = x$ (figur 24.4).

Om $x = f(x)$, så är $x = f(f(x)) = f_2(x)$. Så $x = 0$ och $x = x^*$ är också fixpunkter för funktionen $f_2(x)$. Men när $r > 3$ har funktionen $f_2(x)$ två andra fixpunkter, x_- och x_+ , sådana att $f(x_-) = x_+$ och $f(x_+) = x_-$.

Vi noterar att $f'_2(x) = f'(f(x))f'(x)$ så att $f'_2(x^*) = [f'(x^*)]^2$. Men $f'(x) = r(1-2x)$ och $x^* = 1 - 1/r$. Vilket medför att $f'(x^*) = 2 - r$ och $f'_2(x^*) = (2 - r)^2$. Följaktligen är lutningen på funktionen $f_2(x)$ vid $x = x^*$ sådan att $f'_2(x^*) > 1$ om $r > 3$. Men eftersom $f_2(0) = 0$,



Figur 24.4: Kurvorna $y = f_2(x) = f(f(x))$ och $y = x$ och jämviktstillståndet x^* . (a) $r = 2,75$. (b) $r = 3,4$ och de två andra lösningarna x_- och x_+ av ekvationen $x = f_2(x)$.

$f_2'(0) = r^2 > 1$ och $f_2(1) = 0$ ser vi i figur 24.4b att det nödvändigtvis finns två andra lösningar x_- och x_+ till ekvationen $x = f_2(x)$, med $0 < x_- < x^*$ och $x^* < x_+ < 1$. Ett annat sätt att komma fram till samma slutsats är att lösa ekvationen $x = f_2(x)$, som är en fjärdegradsekvation med två kända rötter: $x = 0$ och $x = x^*$. De två andra lösningarna x_- och x_+ är rötterna till polynomet

$$x^2 - \frac{1+r}{r}x + \frac{1+r}{r^2} = 0. \quad (24.3)$$

De är reella om diskriminanten är positiv, dvs. om $r > 3$. Eftersom $f_2(f(x_-)) = f(f(f(x_-))) = f(f_2(x_-)) = f(x_-)$ är punkten $f(x_-)$ också en fixpunkt för $f_2(x)$. Men $f(x_-) \neq x_-$ eftersom x_- inte är en fixpunkt för $f(x)$. Och $f(x_-) \neq x^*$, annars skulle vi ha $x_- = f(f(x_-)) = f(x^*) = x^*$. Eftersom $f(x_-) \neq 0$ drar vi slutsatsen att $f(x_-) = x_+$. På samma sätt är $f(x_+) = x_-$.

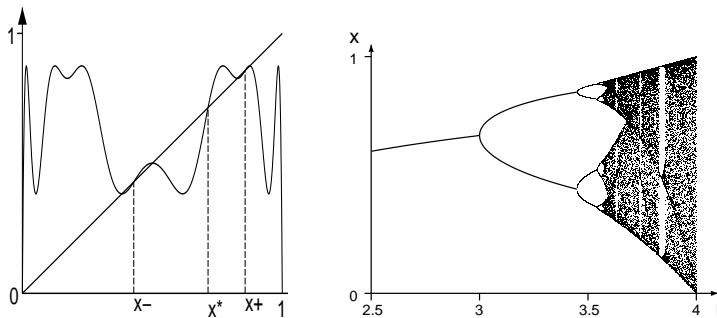
För $r > 3$ ser vi alltså att om till exempel $x_0 = x_-$, så är $x_1 = x_+$, $x_2 = x_-$, $x_3 = x_+$, osv. Man kan också visa att för nästan varje begynnelsevillkor $0 < x_0 < 1$ går talföljden x_n , då $n \rightarrow +\infty$, mot den 2-periodiska cykeln: $x_-, x_+, x_-, x_+,$ osv. (figur 24.3b och 24.4b). Denna cykel förblir stabil så länge r är under det kritiska värdet $r_1 = 1 + \sqrt{6} \approx 3,449$, där $f_2'(x_-) = -1$.

Med hjälp av (24.3) ser vi faktiskt att

$$\begin{aligned} f_2'(x_-) &= f'(f(x_-)) f'(x_-) = f'(x_+) f'(x_-) = r^2 (1 - 2x_+) (1 - 2x_-) \\ &= r^2 [1 - 2(x_+ + x_-) + 4x_+x_-] = r^2 \left[1 - 2 \frac{1+r}{r} + 4 \frac{1+r}{r^2} \right] = -r^2 + 2r + 4. \end{aligned}$$

Så $f_2'(x_-) = -1$ om $-r^2 + 2r + 5 = 0$ och i synnerhet om $r = 1 + \sqrt{6}$.

För $r_1 < r < r_2$ blir en cykel med period 4 stabil: Fyra nya fixpunkter för funktionen $f_4(x) = f_2(f_2(x)) = f(f(f(f(x))))$ uppstår (figur 24.5a). För $r_2 < r < r_3$ är det en cykel med längden 8 osv. Talen r_n går mot gränsvärdet $r_\infty \approx 3,570$ när $n \rightarrow +\infty$. Då $r_\infty < r \leq 4$ kan systemet till och med vara kaotiskt! Figur 24.5b visar bifurkationsdiagrammet², vilket ger en uppfattning om dynamikens komplexitet.



Figur 24.5: (a) Kurvan $y = f_4(x)$ när $r = 3,5$ och linjen $y = x$. Förutom x^* , x_+ och x_- finns det fyra andra fasta punkter som inte är lätta att urskilja. (b) Bifurkationsdiagram för ekvation (24.2).

R. M. May avslutade med att betona att även mycket enkla dynamiska system kan ha ett mycket komplicerat beteende. Detta är inte specifikt för ekvationen $x_{n+1} = rx_n(1 - x_n)$. Samma "kaskad av periodför-dubblingar" som leder till kaos förekommer för andra ekvationer med en funktion $f(x)$ som är formad som en "backe" med ett krön. Detta är till exempel fallet med en annan ekvation som används inom populationsbiologin: $x_{n+1} = x_n \exp(r(1 - x_n))$.

²Detta diagram erhöles genom att för varje givet värde på r plotta punkterna med koordinaterna (r, x_{200}) , (r, x_{201}) , ..., (r, x_{220}) , där $x_{n+1} = f(x_n)$ och $x_0 = 0,1$. Om x_n går mot ett jämviktstillstånd ser vi bara en punkt i diagrammet. Om x_n går mot en cykel med period 2 ser vi två punkter osv.

Den här studien visar att man inte bör bli förvånad över att många datamängder om populationsdynamik är svåra att analysera. Modellen visar också att skillnaden mellan deterministiska och stokastiska modeller inte är så tydlig som man tidigare trott; även med en enkel deterministisk modell kan det vara omöjligt att göra långsiktiga prognoser om parametrarna befinner sig i den kaotiska regimen.

År 1979 valdes May in i *Royal Society*. Från 1988 till 1995 var han professor vid Oxfords universitet och *Imperial College* i London. Mellan 1995 och 2000 var han vetenskaplig chefsrådgivare till den brittiska regeringen. År 1996 fick han Crafoordpriset för sin banbrytande ekologiska forskning rörande teoretisk analys av populationers, samhällens och ekosystems dynamik. Från ekologin vände han sig till epidemiologi och immunologi och publicerade två böcker: *Infektionssjukdomar hos människor* (1991, tillsammans med Roy Anderson) och *Virusdynamik, de matematiska grunderna för immunologi och virologi* (2000, tillsammans med Martin Nowak). I den sistnämnda boken analyseras samspelet mellan immunförsvarets celler och hiv (det virus som orsakar aids) som ett slags rovdjur-bytesdjurssystem (se kapitel 13). Från 2000 till 2005 var May ordförande för *Royal Society*. Han blev adlad 1996 och blev en livstidspärer 2001. Han avled 2020.

Ytterligare läsning

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

Kapitel 25

Kinas ettbarnspolitik (1980)

År 1980 beräknade Song Jian och hans medarbetare, som hade specialiserat sig på kontrollteori tillämpad på lufterumsteknik, att om födelsetalet i Kina förblev på samma nivå som 1980 skulle befolkningen nå över två miljarder under 2000-talet. Deras resultat, som byggde på en matematisk modell med åldersstruktur, bidrog till regeringens beslut att övergå till en ettbarnspolitik.

Song Jian¹ föddes 1931 i Rongcheng i den kinesiska provinsen Shandong. Under 1950-talet studerade han i Sovjetunionen vid Bauman Moskvas statliga tekniska universitet och vid matematiska och mekaniska institutionen vid Moskvas statliga universitet. Han återvände sedan till Kina och blev chef för kontoret för cybernetisk forskning vid matematiska institutet vid den kinesiska vetenskapsakademien. Han var specialist på tillämpningen av kontrollteori för styrning av missiler. Han arbetade också för det sjunde ministeriet för maskinbyggande, som senare bytte namn till ministeriet för luft- och rymdteknik. År 1978 började han fokusera på kopplingarna mellan kontrollteori och demografi.



Figur 25.1: Song Jian

För att förstå sammanhanget kring Song Jian's arbete om populationsdynamik bör man först få en uppfattning om vad "kontrollteori" är. Det är en studie av dynamiska system vars beteende beror på vissa parametrar som kan ändras med tiden för att optimera ett visst kriterium.

¹Song är ett efternamn. Det skrivs alltid först på kinesiska.

Denna teori hade särskilt utvecklats i samband med rymdprogrammen i USA och Sovjetunionen. Ingenjörer var nämligen tvungna att "kontrollera" rymdfärjornas bana för att föra satelliterna till sin omloppsbana runt jorden. Men tillämpningarna var inte begränsade till fysiska eller tekniska problem. Födelsekontrollpolitik kan också betraktas som ett slags problem med optimal kontroll i matematisk mening.

Man bör också nämna uppsatsen *Tillväxtens gränser: en rapport utarbetad för Romklubbens projekt "Mänsklighetens situation"*, som publicerades 1972 och som skrevs av en grupp från teknikinstitutet i Massachusetts (M.I.T.). Studien baserades på en matematisk modell av världens ekonomiska tillväxt som tog hänsyn till naturresurser, befolkningsstorlek och föroreningar. I rapporten antydde att världsekonomin var på väg mot en katastrof på grund av utarmning av icke förnybara resurser, brist på mat för befolkningen eller kraftiga föroreningar. En frivillig begränsning av antalet födselar var en av de föreslagna lösningarna. Sammanfattningsvis var det ett slags modern version av Malthus teser. Rapporten fick ett stort eko i västvärlden under 1970-talet.

Sedan Folkrepubliken grundades 1949 har det kinesiska födelsetalet varit mycket högt utom under det katastrofala "Stora språnget". I mitten av 1970-talet höll Kina långsamt på att återhämta sig från kulturrevolutionen. Familjeplanering uppmanade kvinnor att skjuta upp födselar, förlänga tiden mellan två på varandra följande födselar och få färre barn. Deng Xiaoping, som blev den nya ledaren efter Mao Zedongs död 1976, inledde 1978 politiken för "fyra moderniseringar": jordbruk, industri, vetenskap och teknik samt nationellt försvar. Den kinesiska befolkningens storlek och tillväxt uppfattades då som viktiga hinder för dessa moderniseringar. Forskare som hittills hade arbetat med militära tillämpningar uppmuntrades att hitta lösningar på detta svåra problem.

Mot denna bakgrund åkte Song Jian 1978 till Helsingfors för att delta i en kongress för Internationella federationen för automatisk kontroll. Där noterade han att vissa forskare i Europa hade försökt tillämpa kontrollteorin på befolkningsproblem med tanken att en strikt födelsekontroll till slut skulle kunna förhindra de katastrofer som aviserades i rapporten om *Tillväxtens gränser*. När han återvände till Kina bildade han ett litet team med sin kollega Yu Jingyuan och dataexperten Li Guangyuan för att tillämpa denna typ av matematisk modellering på uppgifter om den kinesiska befolkningen. Vid den tiden var den vetenskapliga kommunikationen mellan Kina och resten av världen knapphändig. Teamet omarbetade ekvationerna som beskriver utvecklingen av en befolknings åldersstruktur på samma sätt som Lotka och McKendrick hade gjort (se kapitlen 10 och 16). Med hjälp av en kontinuerlig

tidsmodell kan man definiera

- $P(x, t)$ befolkningen i åldern x vid tidpunkten t ;
- $m(x)$ dödligheten vid ålder x ;
- $P_0(x)$ befolkningens åldersstruktur vid tiden $t = 0$;
- $b(t)$ Kvinnornas totala fertilitet vid tid t , dvs. det genomsnittliga antal barn som en kvinna skulle få under sitt liv om den åldersspecifika fertiliteten förblev densamma som vid tiden t ;
- f andelen födda kvinnor;
- $h(x)$ sannolikhetsfördelningen för moderns ålder när ett barn föds, med

$$\int_0^{+\infty} h(x) dx = 1.$$

Med dessa noteringar och hypoteser kan åldersstrukturens utveckling modelleras med hjälp av den partiella differentialekvationen

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

med initialvillkoret $P(x, 0) = P_0(x)$ och randvillkoret

$$P(0, t) = b(t) f \int_0^{+\infty} h(x) P(x, t) dx,$$

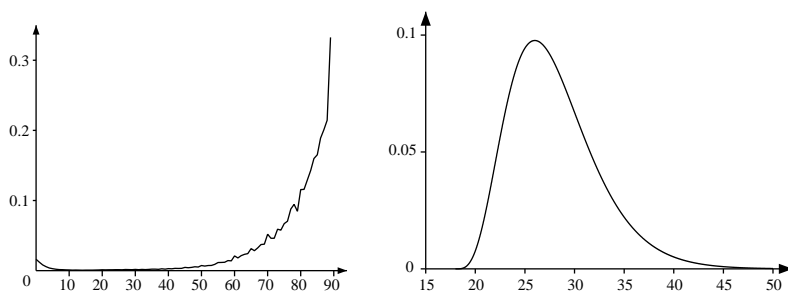
där $b(t)$ är den parameter som ska kontrolleras. Om kvinnornas totala fertilitet är konstant och ligger över det kritiska tröskelvärdet

$$b^* = 1 / \left[f \int_0^{+\infty} h(x) e^{-\int_0^x m(y) dy} dx \right]$$

ökar befolkningen exponentiellt. Detta kriterium liknar det som Lotka fick fram med formeln (10.2). Song Jians grupp tog även hänsyn till den tidsdiskreta versionen av modellen, som liknar Leslies modell (se kapitel 21). Låt $P_{k,n}$ vara befolkningen i åldern k år n . Introducera på samma sätt m_k , b_n och h_k . Då gäller

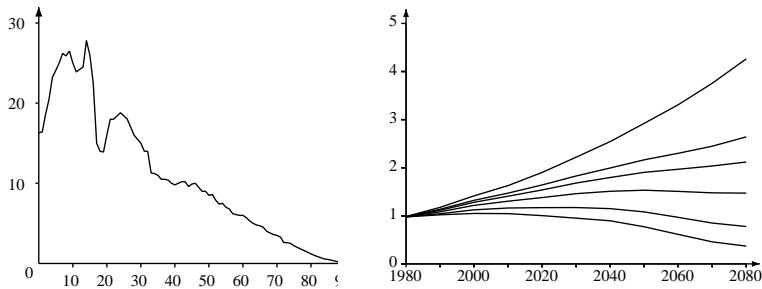
$$P_{k+1,n+1} = (1 - m_k) P_{k,n}, \quad P_{0,n+1} = b_n f \sum_{k \geq 0} h_k P_{k,n}.$$

Genom att känna till dödligheten m_k (figur 25.2a), andelen födda kvinnor $f \approx 0,487$, åldersfördelningen för mödrar h_k (figur 25.2b), det initiala villkoret $P_{k,0}$ som är befolkningens åldersstruktur 1978 (figur 25.3a) och varierande total fertilitet b (som antas vara konstant under varje simulering), kunde Song Jian's team göra demografiska prognoser för sitt land med en tidshorisont på hundra år, från 1980 till 2080 (figur 25.3b). Med tanke på de tusentals adderingar och multiplikationer som krävdes (årtalet n varierar mellan 0 och 100 år, åldern k mellan 0 och 90 år) var en dator nödvändig. Vid den tiden i Kina var det få människor som hade tillgång till sådan utrustning, utom de som arbetade för militären. Song Jian, en ledande expert på missilstyrning, var en av dem.



Figur 25.2: (a) Dödlighet (per år) som en funktion av åldern 1978. (b) Utjämnad form av fertiliteten (per år) som en funktion av åldern 1978.

Enligt prognoserna skulle befolkningen öka från 980 miljoner 1980 till 2,12 miljarder 2080, även om Kina bibehåller sin fertilitet från 1978 på 2,3 barn per kvinna, vilket ligger strax över den kritiska tröskeln som beräknas vara 2,19 barn per kvinna. Men Kina använde redan nästan all mark som kunde användas för jordbruk. Landet hade till och med en tendens att förlora en del av denna mark på grund av ökenspridning och urbanisering. Hur ska man försörja en sådan befolkning om utvecklingen av jordbrukets avkastning inte är tillräcklig? Det är samma fråga som Malthus hade funderat på två århundraden tidigare. Med 1975 års fruktsamhet på $b = 3,0$ kan befolkningen till och med nå 4,26 miljarder år 2080. Med $b = 2,0$ skulle befolkningen nå ett maximum på 1,53 miljarder runt år 2050 innan den börjar minska något. Med $b = 1,5$ skulle en högsta nivå på 1,17 miljarder människor uppnås runt 2030. Med $b = 1,0$ skulle den högsta siffran bara vara 1,05 miljarder och uppnås



Figur 25.3: (a) Ålderspyramiden 1978. Horisontell axel: ålder. Vertikal axel: befolkning (i miljoner). (b) Demografiska prognoser (i miljarder) enligt olika hypoteser om det genomsnittliga antalet barn per kvinna. Från botten till toppen: $b = 1,0; 1,5; 2,0; 2,3; 2,5; 3,0$.

runt år 2000. Under detta antagande skulle befolkningen återgå till 1978 års nivå först år 2025.

Den mest överraskande delen av detta arbete var dess praktiska konsekvenser, som faktiskt är av oöverträffad betydelse i den matematiska befolkningsdynamikens historia. Li Guangyuan visade resultaten av gruppens simuleringar i december 1979 under ett symposium om befolkning i Chengdu, Sichuanprovinsen². I januari 1980 publicerade Song Jian, Yu Jingyuan och Li Guangyuan dessa resultat i en kinesisk ekonomisk tidskrift och förespråkade för övrigt en ettbarnspolitik. De skickade också sin artikel – *En rapport om kvantitativ forskning om frågan om Kinas befolkningsutveckling* – till Kinas främste vetenskapsman Qian Xuesen, som vidarebefordrade den med rekommendation till chefen för administrationen för födelseplanering. Resultaten från Song Jians grupp gjorde ett djupt intryck på de flesta politiska ledare. Dessa var redan övertygade om nödvändigheten av en ökad födelsekontroll trots vad Marx hade skrivit (se kapitel 5) men tvekade fortfarande om kontrollnivån. I februari 1980 fastställde statsrådet och partiets centralkommitté ett mål för den kinesiska befolkningen på 1,2 miljarder för horisonten 2000. I mars 1980 offentliggjordes resultaten från Song Jians grupp i *Folkets Dagblad*. I april undersökte en kommission bestående av politiska ledare och befolkningspecialister de miljömässiga

²Här och nedan sammanfattar vi Susan Greenhalghs detaljerade redogörelse [1,2].

och ekonomiska konsekvenserna av befolkningstillväxten och drog slutsatsen att en ettbarnspolitik var nödvändig för att nå det mål som Deng Xiaoping satt upp för inkomsten per capita år 2000. Politiken blev officiell i september samma år och ett öppet brev som förklarade den för befolkningen publicerades på första sidan i *Folkets Dagblad*.

År 1983 kommer det fortfarande att finnas många obehöriga födslar. Det beslutades att en medlem av varje par med redan två barn skulle steriliseras och att varje förbjuden graviditet skulle avbrytas. Från och med 1984 fick dock landsbygdspär med bara en dotter tillåtelse att få ett andra barn. Ettbarnspolitiken gällde fram till 2015, men under de senaste åren har vissa justeringar gjorts: Om både mannen och kvinnan i ett par är ensam barn kan de få två barn. De repressiva åtgärderna mot par som hade fler än ett barn var hårda: Statsanställda kunde förlora sina jobb, man måste betala dyra böter för att få de administrativa pappren för att skriva in ett andra barn osv. Sammanfattningsvis är det svårt att hitta något annat exempel i den matematiska modelleringens historia med en så stark social inverkan. Song Jian och hans medarbetares arbete var naturligtvis bara en av de faktorer som ledde till beslutet att införa ettbarnspolitiken, men den verkar ha spelat en viktig roll.

Liksom i tidigare kapitel kan den matematiska modelleringens roll vara ett problem. Med utgångspunkt i en verklig situation byggs en modell upp. Den kan analyseras matematiskt eller simuleras med en dator. Man kan sedan förstå hur modellen beter sig när vissa parametrar varierar. Matematiken säger dock inte om modellen är en trogen bild av verkligheten. Vissa mycket viktiga aspekter kan ha försummats. Vissa modeller innehåller också en målfunktion, till exempel att hålla den kinesiska befolkningen under 1,2 miljarder människor. Matematiken säger inte om detta mål var lämpligt³.

År 1980 var Song Jian också medförfattare till den nya upplagan av boken *Ingenjörscybernetik* av Qian Xuesen, "fadern" till det kinesiska rymdprogrammet. Han innehade sedan olika politiska befattningar på hög nivå: vice minister och chefsingenjör vid ministeriet för luft- och rymdteknik (1981–1984), medlem av det kinesiska kommunistpartiets centralkommitté (1982–2002), ordförande för den statliga kommissionen för vetenskap och teknik (1985–1998), statsråd (1986–1998) osv. Han publicerade också två andra böcker som har översatts till engelska: *Befolkningskontroll i Kina* (1985, med Tuan Chi-Hsien och Yu Jingyuan) och *Kontroll av befolkningssystemet* (1988,

³Befolkningen år 2000 beräknades vara 1,264 miljarder. Inkomsten per capita har ökat ungefär från \$200 till \$1 000 mellan 1980 och 2000. Samtidigt har könskvoten blivit extremt snedfördelad till förmån för pojkar, främst på grund av könsslekta aborter.

med Yu Jingyuan). I dessa böcker utvecklas teorin om optimal kontroll som tillämpas på befolkningsdynamik. Song Jian valdes 1991 in i Kinas vetenskapsakademi och 1994 i ingenjörsakademin, som han var ordförande för från 1998 till 2002.

Ytterligare läsning

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. (översättare: Eklöf, M.): *Tillväxtens gränser : en rapport utarbetad för Romklubbens projekt "Mänsklighetens situation"*. Bonnier, Stockholm (1972)
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)

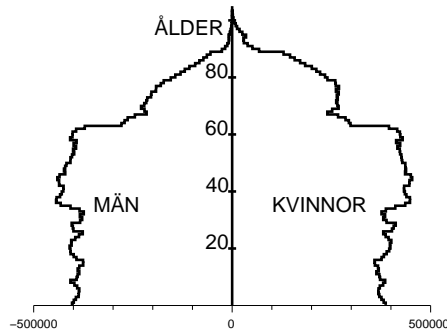
Kapitel 26

Några samtida problem

Detta kapitel ger en kort översikt över några aktuella problem inom matematisk populationsdynamik: befolkningens åldrande inom demografin, nya sjukdomar (aids, sars, vektorburna sjukdomar...) och vaccinationspolitik inom epidemiologin, fiskepolitik inom ekologin och spridning av genetiskt modifierade organismer inom populationsgenetiken. De specialiserade institutioner som arbetar i Frankrike med modellering av dessa problem nämns. Olika aspekter av forskningsarbetet lyfts också fram.

Inom demografin har ett relativt nytt problem dykt upp under de senaste decennierna: befolkningens åldrande. Detta problem är ett bekymmer inte bara i Frankrike (figur 26.1) utan även i många andra europeiska länder och i Japan. Det har viktiga ekonomiska och sociala konsekvenser: pensionssystem, invandringspolitik osv. I Frankrike utvecklas matematiska modeller för att analysera fenomenet åldrande av det nationella institutet för demografiska studier (INED) och det nationella institutet för statistik och ekonomiska studier (INSEE). En av svårigheterna med demografiska prognoser är att födelsetalen kan variera kraftigt över tiden utan att man kan förutse det ens ett decennium i förväg. Detta är särskilt slående om man ser tillbaka på de prognoser som gjordes 1968 för den franska befolkningen 1985: Dessa prognoser kunde inte förutse den minskning av födelsetalen som inträffade under 1970-talet. Det skulle vara intressant att gå igenom alla de prognoser baserade på matematiska modeller som visade sig vara felaktiga, särskilt de som fick ett eko i media. Detta skulle motverka det intryck av "framsteg" som ges i denna bok, ett intryck som kanske redan har förefallit misstänkt för läsaren efter att ha läst kapitlet om den kinesiska ettbarnspolitiken. När det gäller det sistnämnda ämnet är ett nytt problem nu aktuellt: Hur man kan mildra politiken för att undvika det snabba åldrande som förväntas under de närmaste decennierna. Återigen bidrar matematiska modeller till debatten.

Inom epidemiologin är utvecklingen av aidsepidemin särskilt slående bland de nya problem som har uppstått globalt under de senaste två decennierna. I vissa modeller försöker man gissa hur epidemin kommer att



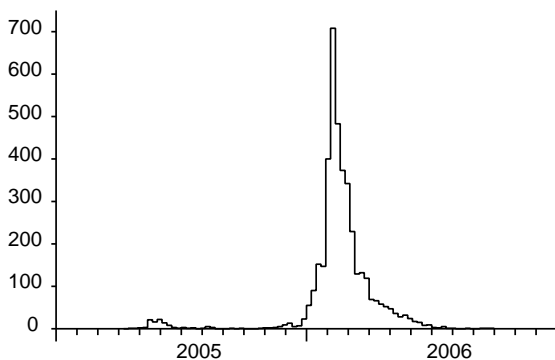
Figur 26.1: Ålderspyramid för den franska befolkningen den 1 januari 2010.
Källa: www.insee.fr.

utvecklas i de länder som nyligen smittats, t. ex. Ryssland, Indien och Kina. Det är svårt att förutsäga om epidemin kommer att avta som i Västeuropa och Nordamerika eller om den kommer att nå en stor andel av befolkningen som i vissa länder söder om Sahara. Andra nya sjukdomar som Ebola i Afrika, West Nile-feber i Nordamerika, sars ("svår akut respiratorisk sjukdom"), fågelinfluensa, chikungunya eller H1N1-influensa har alla undersökts med matematiska modeller, även om de inte har varit särskilt framgångsrika.

När det gäller sars var en svårighet med modelleringen att epidemin var relativt begränsad inom varje land, men kunde spridas mycket snabbt från land till land (Hongkong och Kina, Singapore, Kanada...). Man kunde inte bortse från att epidemikurvorna i varje nytt område var slumpmässiga. Som vi såg i kapitlen 16 och 22 är stokastiska modeller vanligtvis svårare att hantera.

När det gäller chikungunya-epidemin som inträffade mellan 2005 och 2006 på ön Réunion (ett franskt utomeuropeiskt territorium i Indiska oceanen) har vissa modeller inspirerats av Ross modell för malaria (se kapitel 12), eftersom de två sjukdomarna överförs av myggor. En viktig aspekt att ta hänsyn till var säsongens inverkan. Myggpopulationen minskar nämligen under den sydliga vintern, vilket minskar överföringen av sjukdomen. Detta framgår av figur 26.2, som visar antalet nya fall som rapporteras varje vecka av ett litet nätverk av ett trettiotal allmänläkare som bara täcker en bråkdel av öns befolkning. Nätverket upptäckte inga nya fall under flera veckor i september och oktober 2005, men överföringen av sjukdomen fortsatte ändå. Matematiska modeller av epidemin har utvecklats vid INSERM (Nationella institutet för hälsa och medicinsk forskning) och IRD (Tropiska forskningsinstitutet). Trots dessa modeller kunde ingen förutse att epidemin inte skulle dö ut före slutet av sydvintern 2005, när den bara hade smittat

några tusen personer. Slutligen blev nästan en tredjedel av öns befolkning smittad, dvs. cirka 266 000 personer. Detta visar, om än nödvändigt, att det kan vara ganska svårt att förutsäga epidemiernas framtid och att det inte är så lätt att under de första dagarna av en epidemi skilja på om den kommer att bli en mindre eller större epidemi. En parallell kan dras till väderprognoser. Denna typ av prognoser bygger numera på intensiva datorsimuleringar av komplicerade matematiska modeller av havet och atmosfären. Trots detta är prognoser som sträcker sig längre än några dagar inte tillförlitliga.



Figur 26.2: Chikungunya-epidemin på ön Réunion 2005–2006. Antal nya fall som rapporteras per vecka av ett litet nätverk av läkare som en funktion av tiden. Den första lilla toppen nåddes i maj 2005, den andra stora toppen i februari 2006. Siffrorna i denna figur måste multipliceras med cirka 67 för att få fram epidemins verkliga omfattning. Källa: www.invs.sante.fr.

Ur en mer teoretisk synvinkel väckte chikungunya-epidemin frågan om hur man ska anpassa begreppet grundläggande reproduktionsantal \mathcal{R}_0 i modeller som utgår från att miljön har säsongsbetingade (t. ex. periodiska) fluktuationer. Anpassningen är inte så okomplicerad och detta väcker viss oro för hur parametern \mathcal{R}_0 har använts för andra epidemier som påverkas av säsongsvariationer, t. ex. influensapandemin H1N1 2009.

Ett annat problem som blir alltmer oroande och som modellerare har försökt analysera är läkemedelsresistens (antibiotika, malariamediciner osv.). Inom epidemiologin är den sedan Daniel Bernoulli och d'Alembert återkommande frågan om hur man skall balansera kostnader och fördelar

när injektionen av ett vaccin medför en potentiell risk fortfarande föremål för kontroverser och kan förbli det eftersom riskkänsligheten förändras. Efter vissa antydningar om att vaccinet mot hepatit B skulle kunna orsaka allvarliga komplikationer i ett mycket litet antal fall stoppade det franska hälsoministeriet 1998 sin vaccinationskampanj i skolor, även om risken föreföll försumbar jämfört med risken att dö efter att ha smittats av hepatit B-viruset.

Inom ekologin är studiet av fiskpopulationernas dynamik fortfarande problematiskt. Trots detta är det meningen att den skall tjäna som vetenskaplig grund för val av fiskekvoter och andra restriktioner. Överfisket av ansjovis i Biscayabukten och överfisket av röd tonfisk i Medelhavet är bara två exempel från senare tid. Eftersom uppskattningen av fiskbeståndet ofta är opålitlig måste modeller som använder sådana uppgifter betraktas med försiktighet. I Frankrike utförs denna typ av undersökningar huvudsakligen av forskningsinstitutet för utnyttjande av havet (IFREMER). Vissa matematiska modeller har också spelat en roll i Internationella valfångstkommissionens tidigare beslut.

Inom populationsgenetiken är spridningen av genetiskt modifierade organismer också ett kontroversiellt ämne som vissa forskare har försökt studera med hjälp av "reaktionsdiffusionsmodeller" inspirerade av Fisher (se kapitel 20). Detta är ett område som tillhör det nationella institutet för jordbruksforskning (INRAE).

På den mer teoretiska sidan av forskningen kan man nämna följande:

- arbeten om partiella differentialekvationer som diffusionsekvationer (se kapitel 20) eller åldersstrukturerade ekvationer (se kapitel 16);
- arbeten om stokastiska modeller med eller utan rumslig dimension (se kapitlen 16 och 22), inklusive arbeten om slumpmässiga nätverk som modellerar spridningen av epidemier och arbeten som söker deterministiska approximationer.

Denna typ av forskning utförs huvudsakligen av tillämpade matematiker. På senare år har flera masterkurser i matematisk biologi införts vid franska universitet och andra högre utbildningsinstitutioner.

I likhet med andra vetenskapliga områden är den matematiska studien av populationsdynamik huvudsakligen organiserad genom:

- "lärdasamfund": Nederländska sällskapet för teoretisk biologi (sedan 1970), Sällskapet för matematisk biologi (1973), Franskspråkiga sällskapet för teoretisk biologi (1985), Kinesiska sällskapet för matematisk biologi (1985), Japanska sällskapet för matematisk biologi

- (1989), Europeiska sällskapet för matematisk och teoretisk biologi (1991), Latinamerikanska sällskapet för matematisk biologi (2002) osv.
- specialiserade tidskrifter: *Acta Biotheoretica* (sedan 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) osv.
 - konferenser (Årsmöte för Sällskapet för matematisk biologi, Beräkningsmässig och matematisk populationsdynamik, Europeiska konferensen om matematisk och teoretisk biologi, osv.).

Endast de delar som uttryckligen hävdar att de befinner sig i gränssnittet mellan matematik och dess tillämpningar på populationsdynamik har nämnts. Men för varje särskilt område (demografi, ekologi, populationsgenetik, epidemiologi osv.) kan man hitta liknande element med en varierande dos av matematisk modellering.

Avslutningsvis uppmanas den intresserade läsaren att ta en titt på de originalartiklar som finns tillgängliga på webben. Adresserna finns i referenserna i slutet av varje kapitel. Som Ronald Fisher en gång skrev om Mendel:

”Vetenskapshistorien har lidit mycket av att lärarna använt sig av andrahandsmaterial och att de omständigheter och den intellektuella atmosfär under vilka de stora upptäckterna gjordes har utplånats. En förstahandsstudie är alltid lärorik och ofta full av överraskningar.”

Ytterligare läsning

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/

Figurer

- Sid 5. Porträtt av Thomas Murray (ca 1687) i Royal Society i Londons ägo. Chapman, S.: Edmond Halley, F.R.S. 1656–1742. *Notes Rec. R. Soc. Lond.* 12, 168–174 (1957) © The Royal Society.
- Sid 11. Porträtt av Emanuel Handmann (1753) i konstmuseets ägo i Basel. *Leonhard Euler 1707–1783, Beiträge zu Leben und Werk.* Birkhäuser, Basel (1983)
- Sid 16. Porträtt som en gång fanns i Peterskyrkan och som förmodligen förstördes under slaget om Berlin 1945. Reimer, K.F.: Johann Peter Süßmilch, seine Abstammung und Biographie. *Arch. soz. Hyg. Demogr.* 7, 20–28 (1932)
- Sid 22. Porträtt av Johann Niclaus Gooth (ca. 1750–1755) i naturhistoriska museets ägo i Basel. Speiser, D.: *Die Werke von Daniel Bernoulli*, Band 2. Birkhäuser, Basel (1982)
- Sid 29. Porträtt av Maurice Quentin de La Tour (1753) i Louvren i Paris.
- Sid 32. Porträtt av John Linnell (1833) i Haileybury Colleges ägo, England. Habakkuk, H.J.: Robert Malthus, F.R.S. (1766–1834). *Notes Rec. R. Soc. Lond.* 14, 99–108 (1959)
- Sid 36. Gravyr av Flameng (1850). Quetelet, A.: Pierre-François Verhulst. *Annu. Acad. R. Sci. Lett. B.-Arts Belg.* 16, 97–124 (1850)
- Sid 41. Heyde, C.C., Seneta, E.: *I. J. Bienaymé, Statistical Theory Anticipated.* Springer (1977) © Académie des sciences, Institut de France.
- Sid 41. Brun, J., Robinet, A. (éd.): *A. Cournot, études pour le centenaire de sa mort.* Economica / Vrin, Paris (1978)
- Sid 45. Bateson, W.: *Mendel's Principles of Heredity.* Cambridge University Press (1913)
- Sid 50. Pearson, K.: *The Life, Letters, and Labors of Francis Galton*, vol. 1. Cambridge University Press (1914)
- Sid 50. Porträtt av Watson i biblioteket på *Trinity College*, universitetet i Cambridge. Kendall, D.G.: Branching processes since 1873. *J. Lond. Math. Soc.* 41, 385–406 (1966)
- Sid 56. Alfred J. Lotka Dokument. © Princeton University Library.
- Sid 60. Titchmarsh, E. C.: Godfrey Harold Hardy 1877–1947. *Obit. Not. Fellows R. Soc.* 6, 446–461 (1949)
- Sid 63. Stern, C.: Wilhelm Weinberg. *Genetics* 47, 1–5 (1962)
- Sid 66. G.H.F.N.: Sir Ronald Ross 1857–1932. *Obit. Not. Fellows R. Soc.* 1, 108–115 (1933) © The Royal Society.

- Sid 75. Whittaker, E.T.: Vito Volterra 1860–1940. *Obit. Not. Fellows R. Soc.* 3, 690–729 (1941)
- Sid 79. Yates, F., Mather, K.: Ronald Aylmer Fisher, 1890–1962. *Biog. Mem. Fellows R. Soc.* 9, 91–120 (1963) © The Royal Society/Godfrey Argent Studio.
- Sid 83. Yates, F.: George Udny Yule. *Obit. Not. Fellows R. Soc.* 8, 308–323 (1952)
- Sid 91. Heyde, C.C., Seneta, E. (eds.): *Statisticians of the Centuries*. Springer (2001)
- Sid 101. britannica.com/EBchecked/topic/252257/J-B-S-Haldane © Bassano and Vandyk Studios.
- Sid 109. Hill, W.G.: Sewall Wright, 21 December 1889 – 3 March 1988. *Biog. Mem. Fellows R. Soc.* 36, 568–579 (1990) © Llewellyn Studios, Chicago.
- Sid 105. Nybølle, H.C.: Agner Krarup Erlang f. 1. Januar 1878 – d. 3. Februar 1929. *Mat. Tidsskr. B*, 32–36 (1929)
- Sid 105. Nørdlund, N.E.: Johan Frederik Steffensen in memoriam. *Nordisk Mat. Tidsskr.* 10, 105–107 (1962)
- Sid 118. Tikhomirov, V.M.: A.N. Kolmogorov. In: Zdravkovska, S., Duren, P.L. (eds.) *Golden Years of Moscow Mathematics*, 2nd edn., 101–128. American Mathematical Society (2007)
- Sid 118. *I. G. Petrowsky Selected Works Part I*. Gordon and Breach, Amsterdam (1996) © Taylor and Francis Books UK.
- Sid 123. Foto: Denys Kempson. Crowcroft, P.: *Elton's Ecologists, a History of the Bureau of Animal Population*. University of Chicago Press (1991)
- Sid 126. © Geoffrey Grimmett.
- Sid 133. 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.
- Sid 133. Harman, O.: *The Price of Altruism*. W. W. Norton, London (2010)
- Sid 138. © Samuel Schläefli / ETH Zürich.
- Sid 145. Selected works of J. Song. *Science Press*, Beijing (1999) © Song Jian.

Innehåll

Inledning	iii
1 Fibonaccis talföljd (1202)	1
2 Återstående livslängd vid given ålder enligt Halley (1693)	4
3 Euler och den geometriska befolkningstillväxten (1748)	10
4 Daniel Bernoulli och ympning mot smittkoppor (1760)	21
5 Malthus och hindren för den geometriska tillväxten (1798)	32
6 Verhulst och den logistiska ekvationen (1838)	36
7 Bienaymé, Cournot och släktnamnens utdöende (1845)	41
8 Mendel och ärftlighet (1865)	45
9 Galton och Watson och frågan om utdöende släktnamn (1873)	49
10 Lotka och teorin om stabila populationer (1907)	56
11 Hardy-Weinbergs lag (1908)	60
12 Ross och malaria (1911)	65
13 Lotka, Volterra och rovdjurs- och bytesdjurssystemet (1920)	71
14 Fisher och det naturliga urvalet (1922)	78
15 Yule och evolution (1924)	82
16 McKendrick om modellering av epidemier (1926)	90
17 Haldane och mutationer (1927)	100
18 Erlang och Steffensen om utdöendeproblemet (1929)	104
19 Wright och slumpmässig genetisk drift (1931)	109
20 Spridning av gener (1937)	115
21 Leslie-matrisen (1945)	122
22 Perkolation och epidemier (1957)	126
23 Spelteori och evolution (1973)	132
24 Kaotiska populationer (1974)	138
25 Kinas ettbarnspolitik (1980)	145
26 Några samtida problem	152

Den här boken följer populationsdynamikens historia, ett teoretiskt forskningsfält som är nära besläktat med genetik, ekologi, epidemiologi och demografi, och som matematiken har bidragit med viktiga insikter till. Den ger en översikt över några viktiga ämnen, såsom exponentiell tillväxt, utvecklingen av stokastiska modeller och kaotiska populationer, där determinism och slumpmekanismer sammanflätas. Därmed görs nedslag i historien om Euler, om Malthus och om den kinesiska ettbarnspolitiken. Dessutom förklaras historien om Mendels lagar och frågan om utdöende släktnamn samt perkolationsteorin för spridning av epidemier.

Med de senaste framstegen inom maskinöversättning är det inte längre berättigat att ett enda språk ska ha monopol på vetenskaplig litteratur. Den ökande språkliga alienationen vid universiteten kan vändas. Med denna omsorgsfullt reviderade svenska översättning främjar vi denna nya väg.

ISBN : 979-10-343-9008-3



15€