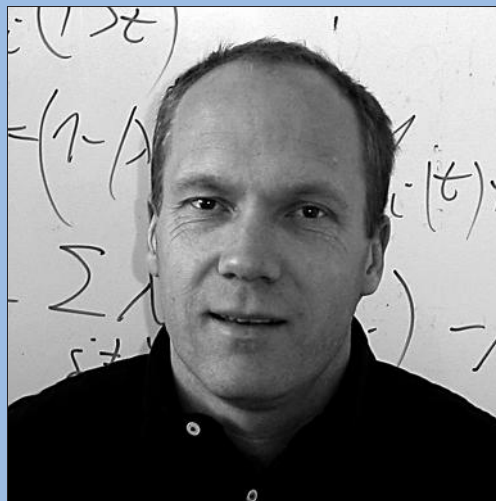


Tom Britton

” *Avvägningen mellan att göra en matematisk modell enkel, för att kunna räkna på den, eller avancerad, för att bättre efterlikna verkligheten, är en svår balansgång.*

Jag började läsa matematik vid högskolan av en anledning som jag tror är vanlig oberoende av ämne, nämligen att jag hade en mycket bra matematiklärare, Ingmar Carlsson vid Södra Latin, som dessutom påpekade att jag hade fallenhet för ämnet och borde läsa vidare. Tack Ingmar! Väl vid högskolan tyckte jag att ämnet var roligt och det gick bra så jag fortsatte mot en magisterexamen. De sista två åren av studierna skulle man profilera sig mot någon del av ämnet.

Att jag valde matematisk statistik, dvs sannolikhetsteori och statistisk teori, var nog inte bara en slump – jag tyckte att denna slumpens matematik var mycket spännande. Det som fascinerade mig mest var för det första vilket kraftigt verktyg slumpen är för att beskriva verkliga fenomen. Inte så att det ofta förekommer ”äkta slump” (jag vet själv inte om det ens finns någon sådan) utan snarare att slump på ett mycket effektivt sätt ”tar hand” om de delar vi inte lyckas fånga upp i vår matematiska modell. Om vi tex planterar ett antal till



synes identiska frön av någon växtsort och odlar upp dem under så identiska former vi kan, så kommer växternas tillväxt inte att bli lika, utan den varierar enligt den välkända normalfördelningen. Att det kommer att bli så (jag har i och för sig aldrig utfört experimentet) beror på att hur noggranna vi än är kommer det att finnas små skillnader i fröna och deras omgivning, och många (små) skillnader brukar tillsammans skapa normalfördelade avvikelser.

Det andra i ämnet som fascinerade mig var de regler och system som den, i första betraktelsen, kaotiska slumpen följer. Att många små variationer tillsammans ger normalfördelningen som i exemplet ovan är en sådan. Denna egenskap kallas centrala gränsvärdessatsen. En annan dylik lag är att medelvärdet av många likartade slumpmässigheter inte längre är särskilt slumpmässigt. Denna ”de stora talens lag” ger sig till känna t ex genom att det i princip räcker att i en opinionsundersökning tillfråga några tusen slumpmässigt utvalda av Sveriges invånare för att med god precision veta vad hela landets invånare tycker.

Nu, när jag forskar och undervisar, inser jag också två nya egenskaper hos sannolikhetssteori och statistisk teori som jag gillar. Det första är att det är väldigt lätt att levandegöra teorin med exempel, något som underlättar och motiverar i undervisning. Det andra är att man som forskare har lätt att finna samarbeten med forskare inom andra ämnesområden. Merparten vetenskapsområden använder sig nämligen av sannolikhetssteori och /eller statistiska modeller, och att samarbeta med forskare i andra ämnesområden är något jag uppskattar alltmer.

Jag är född 1965 och uppvuxen i Stockholm. Jag disputerade vid Stockholms universitet 1996 och var därefter postdoktorstipendiat i Australien ett år och lektor/forskare vid Uppsala universitet i 5 år. Sedan 2003 är jag professor i matematisk statistik vid Stockholms universitet där jag forskar, handleder och undervisar. Min forskning rör sannolikhetssteori och statistiska metoder med tillämpningar inom biologi och medicin. Framför allt har jag arbetat med modeller och statistiska metoder för hur smittsamma sjukdomar sprids i en befolkning och hur man drar slutsatser om växtarters evolutionära träd utifrån DNA-sekvenser från växterna.

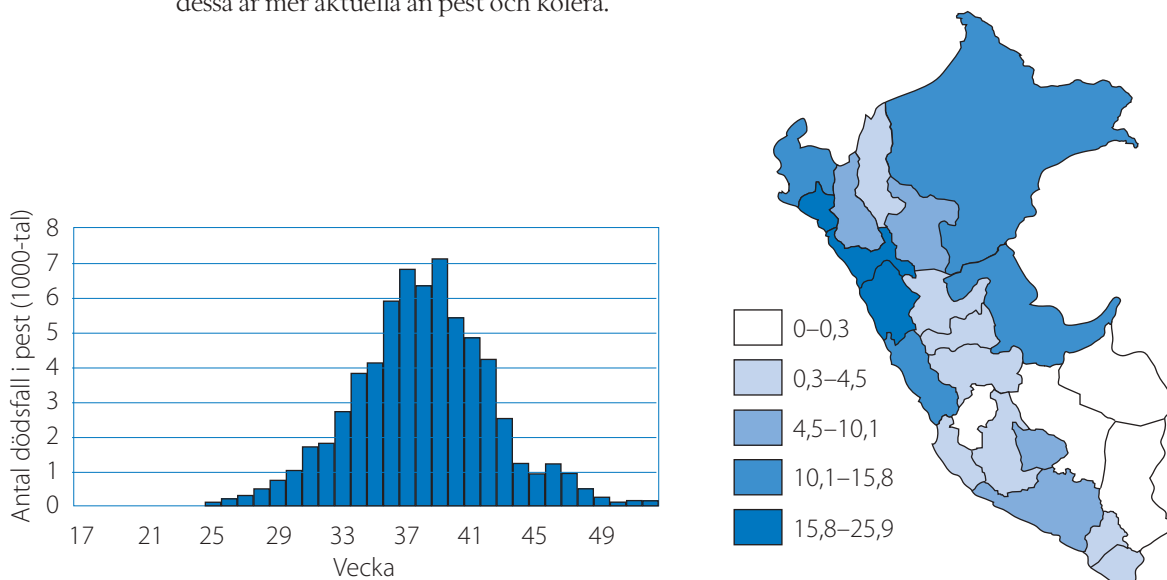
Pest, kolera och matematik

– vad kan matematik och statistik lära oss om smittsamma sjukdomars utbredning?

Detta kapitel tar upp frågeställningar om smittsamma sjukdomars utbredning i en befolkning och hur utbredningen minskar vid vaccinering. Med hjälp av relativt enkla matematiska modeller kommer vi fram till slutsatser om hur stor andel som smittas i en befolkning, givet att vi vet sjukdomens *reproduktionstal* – ett mått på dess smittsamhet. Ur ett statistiskt perspektiv kan man därför skatta reproduktionstalet genom att observera hur många som smittades. Den viktigaste användningen av denna skattning är vid vaccination: om vi vet reproduktionstalet kan vi bestämma hur många som måste vaccineras för att undvika framtida utbrott. Genom att observera hur många som smittades under ett utbrott i en viss befolkning kan vi således skatta hur många som minst måste vaccineras i en liknande befolkning för att denna skall klara sig undan ett utbrott. För att kunna tillgodogöra sig allt i detta kapitel krävs matematik motsvarande gymnasiekompetens. Inga speciella kunskaper om smittsamma sjukdomar förutsätts. Ett visst mått av nyfikenhet skadar inte, något som gäller nästan all matematik.

Pest och kolera är två bland många infektionssjukdomar som har orsakat mycket mänskligt lidande genom tiderna. Pest är numera utrotad medan kolera fortfarande förekommer relativt frekvent i tredje världen. Till vänster i figur 1 ser du data från ett utbrott av pest i London 1665 medan kartan till höger visar antal dödsfall i ett kolerautbrott i Peru 1991.

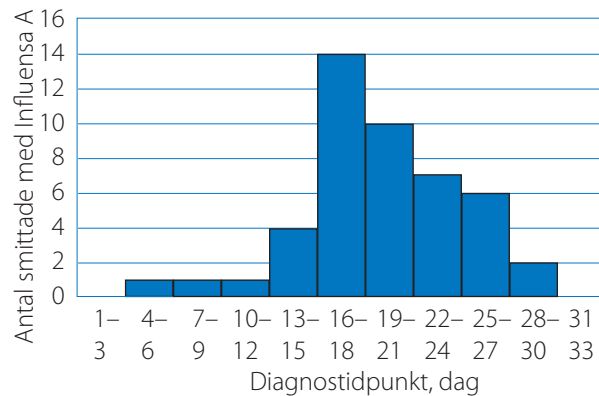
Eftersom dessa och andra sjukdomar drabbar många är det önskvärt att lära sig om utbredningen av smittsamma sjukdomar och hur man kan minska densamma. Jag kommer i kapitlet att beskriva en smittsam sjukdoms utbredning i en befolkning med hjälp av enkla matematiska modeller, och utifrån dessa försöker jag besvara frågor om hur man drar slutsatser för konkreta sjukdomsutbrott. Jag exemplifierar med influensa och mässling eftersom dessa är mer aktuella än pest och kolera.



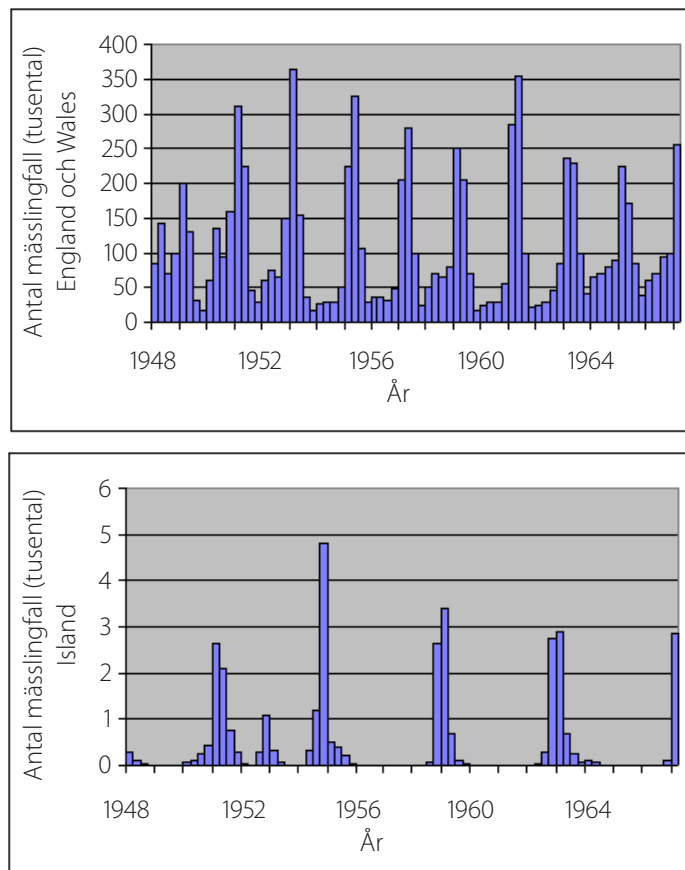
Figur 1. Det vänstra diagrammet visar antalet dödsfall (i tusental) orsakade av pest vecka för vecka i London under ett utbrott 1665–1666. Kartan till höger beskriver antal kolerafall per 1000 invånare i respektive region i Peru under ett kolerautbrott 1991. Källa: Anderson & May respektive WHO.

Figur 2 illustrerar ett typiskt influensautbrott, detta bland anställda och boende på ett ålderdomshem i Rochester, Minnesota, under november/december 1996. Som synes är kurvan inte helt olik den för pest i London 1665 i figur 1 – utbrott av detta slag där många smittas under kort tid varefter utbrottet klingar av brukar kallas epidemi.

Under utbrottet i Minnesota smittades 43 individer av totalt 129 boende och personal, dvs 33%. Man kan ställa ett flertal frågor relaterade till diagrammet. Varför smittades endast 33% och inte alla? Varför växer kurvan så hastigt i början för att sedan dö ut lika snabbt trots att endast en bråkdel smittades? Hur många hade man behövt vaccinera för att undvika utbrottet, eller snarare, hur många bör vaccineras nästa gång en influensaepidemi är på gång för att ett utbrott ska undvikas?



Figur 2. Ett influensautbrott bland boende och personal på ett ålderdomshem i Minnesota 1996. Totala antalet smittade var 43 av 129 möjliga. Källa: Kuhle & Breese, 1998:3, Annals of long-term care.



Figur 3. Diagrammen visar antalet mässlingsfall under perioden 1948–1967, innan vaccinering startades, i England och Wales (övre diagrammet) respektive på Island (nedre diagrammet). Observera att skalorna är olika på de två y-axlarna. Källa: Anderson & May

Vi tar nu ett nytt exempel, som illustreras i figur 3. Denna gång visar de två diagrammen i figuren antalet mässlingsfall under en 20-årsperiod, 1948–1967, i England och Wales, respektive på Island. Även dessa diagram ger upphov till ett flertal naturliga frågor: Varför verkar utbrotten ske periodiskt, dessutom med olika periodlängd (2 respektive 4 år för England och Island)? Varför försvinner mässlingen helt mellan epidemierna på Island medan sjukdomen hela tiden finns kvar i England – ett sk *endemiskt* tillstånd? Hur många måste här vaccineras om man vill få bort mässlingen ur landet?

Vi ska nu se hur matematik, sannolikhets teori och statistik kan användas för att besvara frågeställningar vid smittsamma sjukdomars utbredning, likt dem som nämdes ovan. Vi kommer att förenkla verkligheten med ett flertal antaganden och studera enkla matematiska modeller som beskriver hur en smitta sprider sig i en befolkning. Förhoppningen är att resultaten även gäller i den mer komplicerade verkligheten. Som så ofta är fallet med matematiska modeller visar det sig att samma modeller även kan användas för att beskriva andra förlopp, i detta fall t ex hur kunskap och rykten sprids i en befolkning men även hur datavirus sprids bland datorer. Här koncentrerar vi oss på smittsamma sjukdomar och kallar modellerna epidemimodeller.

Viktiga egenskaper hos smittsamma sjukdomar

Jag vill sätta upp matematiska modeller som beskriver hur en smitta sprids i en befolkning för att kunna få svar på de frågor som formulerades i förra avsnittet och som motiverades av figurerna. Jag fokuserar på sjukdomar som sprids från människa till människa och utesluter därmed infektionssjukdomar som sprids via mat, t ex salmonella, eller via djur, t ex malaria. En smitta kan endast spridas från en person till en annan om den första personen är *smittsam* och den andre är *mottaglig*. Utöver dessa två tillstånd kan man även vara *immun*, dvs man är inte smittsam men kan ej heller smittas. De två möjliga övergångarna är att gå från mottaglig till smittsam (när man smittas), samt att upphöra att vara smittsam och bli immun (när man tillfrisknar). Denna klass av modeller brukar kallas SIR efter engelskans motsvarande tillståndsnamn och övergångar: Susceptible – Infectious – Recovered.

Utöver dessa tre tillstånd, och övergångarna däremellan, är den viktigaste matematiska egenskapen hos en smittsam sjukdom hur stor chans det är att en smittsam individ under sin smittoperiod smittar andra individer. Andra egenskaper, som hur själva smittan sprids, symptom på sjukdomen mm, har bara indirekt betydelse för sjukdomens utbredning. Inkubationstid, latensperiod och infektionsperiodens längd påverkar tidsdynamiken av sjukdomens utbredning men däremot inte utsagor om epidemins storlek eller den endemiska nivån (andelen smittsamma vid endemiskt tillstånd). Om vi tex jämför två sjukdomar där den ena är hälften så smittsam per tidsenhet, men man är smittsam dubbelt så länge, kommer i genomsnitt lika många att smittas av de två sjukdomarna även om förloppet är snabbare för den första sjukdomen.

Somliga sjukdomar har så kallade subkliniska fall, dvs individer som smittas men inte blir smittsamma. Likaså finns det sjukdomar där inte alla blir immuna. Att bortse från dessa fall är dock ingen större inskränkning eftersom de brukar vara sällsynta. Den stora förenklingen jag gör nedan är att anta att alla individer betar sig likadant och har samma förutsättningar att få sjukdomen och sprida den vidare. Konsekvensen blir att vi antar att man smittar varje annan individ med samma sannolikhet, något som förstås inte är sant i verkliga livet eftersom det tex är större chans att smitta familjemedlemmar och nära vänner jämfört med personer man inte känner. Förenklingen gör det dock väsentligt enklare att räkna i den matematiska modellen. Det finns mer avancerade matematiska modeller som tillåter varierande sannolikheter att smitta olika individer beroende på om de ingår i samma familj eller liknande. Även modeller där somliga individer är mer eller mindre mottagliga eller smittsamma förekommer. I en större befolkning bör man dock försöka hålla nere antalet olika sannolikheter i sin modell; att låta alla par av individer ha olika smittosannolikhet skulle göra det praktiskt ogenomförbart att använda modellen.

Avvägningen mellan att göra en matematisk modell enkel, för att kunna räkna på den, eller avancerad, för att bättre efterlikna verkligheten, är en svår balansgång. En matematisk modellerare försöker alltid ta med de "viktigaste" egenskaperna i det hon studerar, men utelämnar mindre viktiga detaljer. Vilka egenskaper som är "viktiga" är inte alltid lätt att avgöra, varför en modell bör granskas kritiskt innan dess slutsatser används för att dra slutsatser om verkligheten.

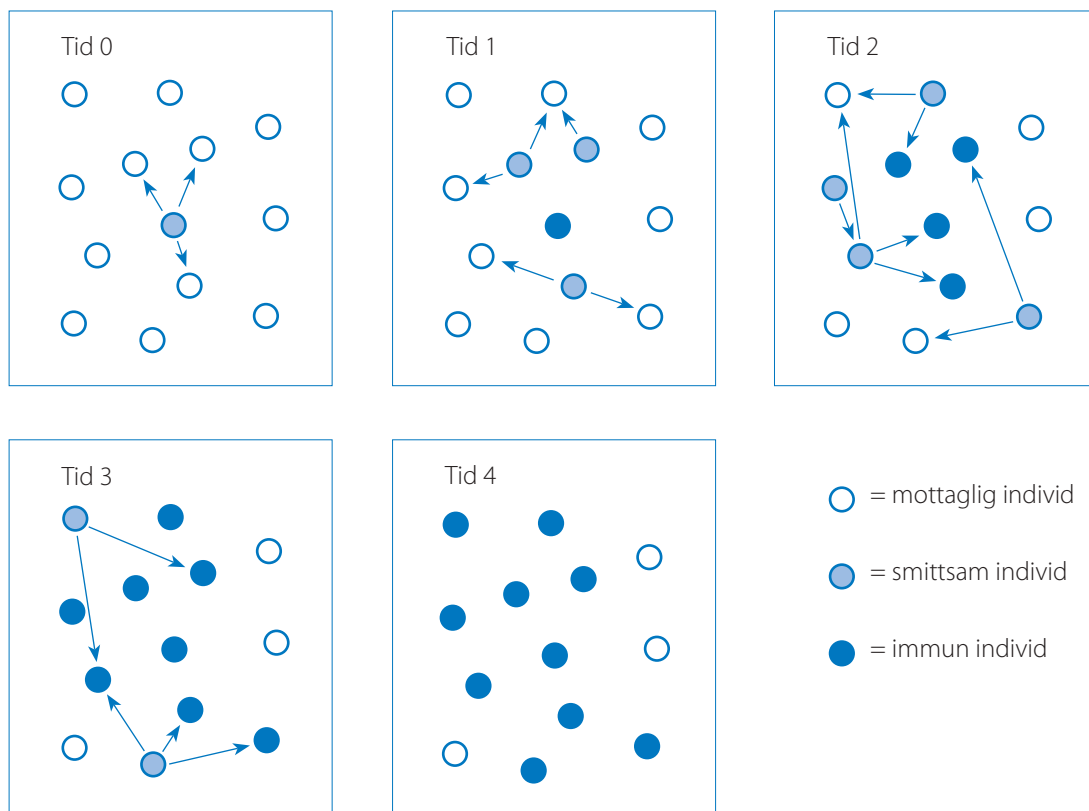
Modellering av epidemier

Jag börjar med en modell som syftar till att beskriva en epidemi, dvs ett kortare sjukdomsutbrott. Eftersom tidsrymden är kort gör jag det förenklande antagandet att befolkningen är konstant under utbrottet och bortser från eventuella födselar, dödsfall, immigration och emigration. Som jag redan nämnt antas vi även att alla individer är lika, dvs att alla har samma sannolikhet att smittas och att alla träffar varandra lika ofta. Jag bortser även från eventuella initialt immuna.

En enkel epidemimodell

1. Initialt är en (eller ett fåtal) individer smittsamma och övriga individer är mottagliga. Om det totalt finns $n + 1$ individer är alltså 1 smittsam och n mottagliga.
2. En smittsam individ har under sin smittoperiod smittsamma kontakter med andra individer slumpmässigt och oberoende, var och en med sannolikheten p . Därefter tillfrisknar individen och blir immun. Mottagliga individer som utsätts för en smittsam kontakt blir i nästa tidssteg smittsamma och beter sig på liknande sätt, dvs de har i sin tur smittsamma kontakter och de kontakter som sker med mottagliga individer resulterar i smittoöverföring, osv.
3. Detta förlopp fortsätter tills dess att inga mottagliga individer smittas. Då stannar epidemin.

Det totala antalet som smittas, förutom initialpersonen, betecknas med $Z(0 \leq Z \leq n)$ och kallas epidemins storlek. I figur 4 illustreras modellen med en serie bilder som beskriver förloppet. Varje bild motsvarar de olika tidsstegen där individer smittas. Pilarna motsvarar en individs smittsamma kontakter, och de kontakter som sker med mottagliga individer resulterar i att de senare smittas.



Figur 4. Figuren visar ett typiskt förlopp i epidemimodellen. Initialt är 1 av totalt 13 individer smittsamma ($n = 12$). Nya individer smittas och smittsamma blir immuna. Efter tid 3 kontaktas inga mottagliga individer varför inga nya smittas och utbrottet stannar av. Vid utbrottets slut har $Z=9$ ytterligare individer smittats.

Som framgår i figur 4 sker efterhand allt fler av kontakterna med individer som redan smittats. Som en följd sjunker antalet nysmittade som en smittsam individ genererar och epidemin stannar till slut när inga nya har smittats. Oftast sker detta utan att alla har smittats. Detta fenomen, att antalet smittsamma först stiger för att sedan avta och till slut bli 0 trots att alla inte smittats, gäller även för sjukdomar i verkliga livet. I figur 2 såg vi ju att antalet nysmittade först tog fart och sedan klingade av innan alla hade smittats. Vi har således hittat en förklaring till detta fenomen, nämligen att allt fler av de smittsamma kontakterna sker med redan smittade och immuna individer.

Sannolikhetsfunktionen för slutstorleken

Den vetgiriga frågar sig hur många som kommer att smittas totalt, dvs vad epidemins slutstorlek blir. Låt oss därför räkna ut sannolikheten för en given slutstorlek, dvs sannolikhetsfunktionen. Denna sannolikhet är förvånansvärt krånglig att räkna ut redan för få individer. Nedan använder jag uttryck av typen $P(\dots) = 0,3$ som uttyds ”sannolikheten för något är 0,3” (” P ” kommer från engelskans ”probability”).

$n = 1$ mottaglig individ

Jag börjar med det allra enklaste fallet, nämligen att det bara finns två människor, en som är smittsam från början och en som är mottaglig. När det bara är en mottaglig individ finns det ju bara två möjligheter: att ingen smittas eller att en smittas. Eftersom sannolikheten för en smittsam kontakt är p , blir sannolikheterna för de två möjliga utfallen:

$$P(0 \text{ smittas}) = 1 - p \quad P(1 \text{ smittas}) = p.$$

$n = 2$ mottagliga individer

Antag nu att det är två mottagliga individer från början (och en smittsam). Att ingen smittas kan bara ske på ett sätt, nämligen att den initialt smittsamme inte smittar någon av de två mottagliga, vilket ger sannolikheten $(1 - p)^2$. För att exakt en individ skall smittas måste först den initialt smittsamme smitta den ene men inte den andre, och sedan får den på så sätt

smittade inte heller smitta ner den andre. Den initialt smittsamme kan ju smitta två olika, så chansen att den smittar en men inte den andre blir $2p(1-p)$. Den andre måste alltså också undgå att smittas från den på så sätt smittade vilket har sannolikhet $(1-p)$. Sannolikheten att exakt en smittas blir således $2p(1-p) \times (1-p)$.

Det lata sättet att räkna ut sannolikheten att bägge smittas är att ta 1 och subtrahera övriga två sannolikheter: summan av alla sannolikheter måste ju bli 1! Ett alternativ är att försöka tänka ut vad sannolikheten skall bli. Man har då möjlighet att kontrollera att man räknat rätt genom att summera de tre sannolikheterna och se att summan blir 1 (blir den inte det finns det anledningen att se över sina beräkningar ...). Att bägge blir smittade kan ske på två olika sätt. Antingen smittas bägge direkt, eller också smittas bara en i första steget, och sedan smittas den andre av den förste. Sannolikheterna för dessa möjligheter blir tillsammans $p^2 + 2p(1-p) \times p$. Vi har kommit fram till följande:

$$P(0 \text{ smittas}) = (1-p)^2$$

$$P(1 \text{ smittas}) = 2p(1-p) \times (1-p)$$

$$P(2 \text{ smittas}) = p^2 + 2p(1-p) \times p.$$

$n=3$ mottagliga individer

Sannolikheterna att 0 respektive 1 individ smittas för fallet $n=3$ mottagliga individer blir analoga med fallet $n=2$: $(1-p)^3$ respektive $3p(1-p)^2 \times (1-p)^2$.

Att två smittas kan ske antingen genom att bägge smittas initialt och ingen av dessa smittar den tredje, eller att först en (av tre) smittas och att denna smittar en av de två kvarvarande mottagliga som i sin tur inte smittar den siste. Sannolikheterna för dessa händelser blir tillsammans: $3p^2(1-p) \times (1-p)^2 + 3p(1-p)^2 \times 2p(1-p) \times (1-p)$. Sannolikheten för den återstående möjligheten kan fås genom subtraktion.

$$P(0 \text{ smittas}) = (1-p)^3$$

$$P(1 \text{ smittas}) = 3p(1-p)^2 \times (1-p)^2$$

$$P(2 \text{ smittas}) = 3p^2(1-p) \times (1-p)^2 + 3p(1-p)^2 \times 2p(1-p) \times (1-p)$$

$$P(3 \text{ smittas}) = 1 - P(0 \text{ smittas}) - P(1 \text{ smittas}) - P(2 \text{ smittas}).$$

Som synes blir det alltmer krångligt att beräkna sannolikheten för de olika slutstorlekarna ju fler mottagliga individer det är från början. När man har några hundra mottagliga individer kan man inte ens beräkna dessa sannolikheter med datorers hjälp. Anledningen därtill är att oerhört stora tal (binomialkoefficienter) skall multipliceras med väldigt små tal, något som ger avrundningsfel även när man räknar med mycket stor noggrannhet. Av denna anledning är det av intresse att finna approximationer som gäller för stora n , något vi gör i nästa del-avsnitt.

Slutstorleken i stora populationer

Modellen innehåller parametern p som är sannolikheten att en smittsam individ smittar en given annan individ i populationen. Om populationstorleken n är stor är det rimligt att anta att p är liten: även om sjukdomen är mycket smittsam är det liten risk att jag stöter ihop med en given individ i samma stad under min smittoperiod. En intressant storhet är därför $R_0 = np$ som brukar kallas *reproduktionstalet*. Tolkningen av R_0 är "genomsnittligt antal smittsamma kontakter man har under sin smittoperiod", vilket är detsamma som antal individer man i genomsnitt smittar under förutsättning att alla är mottagliga. Vi ska nu härleda en ekvation för hur stor andel x som kommer smittas i en stor population, en ekvation som beror på R_0 . Om x är andelen smittade betyder det att $1-x$ är andelen som inte smittas och att nx är antalet som smittas. I serien likheter nedan utnyttjar jag att $p = R_0/n$ är litet. Om man i genomsnitt smittar R_0 individer ($R_0 = 10$ är ett vanligt tal) blir ju chansen p att smitta en enskild individ i en stor population liten. Och, om p är litet gäller approximationen $1-p = 1-R_0/n \approx e^{-R_0/n}$, där e^x är den förhoppningsvis välkända exponentialfunktionen.

Denna approximativa likhet förklaras av Taylorseriens utseende för exponentialfunktionen, något jag inte visar utan som läsaren får ta för givet.

p är sannolikheten att en smittsam individ smittar en given annan individ i populationen.

R_0 är "genomsnittligt antal smittsamma kontakter man har under sin smittoperiod".

$R_0 = np$, brukar kallas *reproduktionstalet*.

epidemins slutstorlek = hur många som kommer att smittas totalt.

$$\begin{aligned}
1 - x &= \text{andel som ej smittas} \\
&= \text{sannolikheten att en enskild individ ej smittas} \\
&= \text{sannolikheten att undgå smittsam kontakt med samtliga } nx \text{ smittade} \\
&\approx (1 - p) \cdot \dots \cdot (1 - p) && nx \text{ faktorer} \\
&= (1 - R_0/n) \cdot \dots \cdot (1 - R_0/n) && nx \text{ faktorer} \\
&\approx e^{-R_0/n} \cdot \dots \cdot e^{-R_0/n} && nx \text{ faktorer} \\
&= e^{-nx \frac{R_0}{n}} = e^{-R_0 x}.
\end{aligned}$$

De två första likheterna är förhoppningsvis inte så svåra att acceptera. Den tredje likheten motiveras av att man undgår en smittsam kontakt om man inte kontaktas av någon av de nx smittade människorna. Den fjärde approximativa likheten förklaras av att chansen för detta i stort sett är chansen att man inte smittas av den först smittade ($=1-p$) multiplicerat med chansen att man inte smittas av den andre smittade ($=1-p$), osv för samtliga nx smittade. Resterande likheter beror på approximationen $1-R_0/n \approx e^{-R_0/n}$ och enkla algebraiska likheter. Vi har alltså med enkel matematik fått fram en balansekvation för x :

$$1 - x = e^{-R_0 x}$$

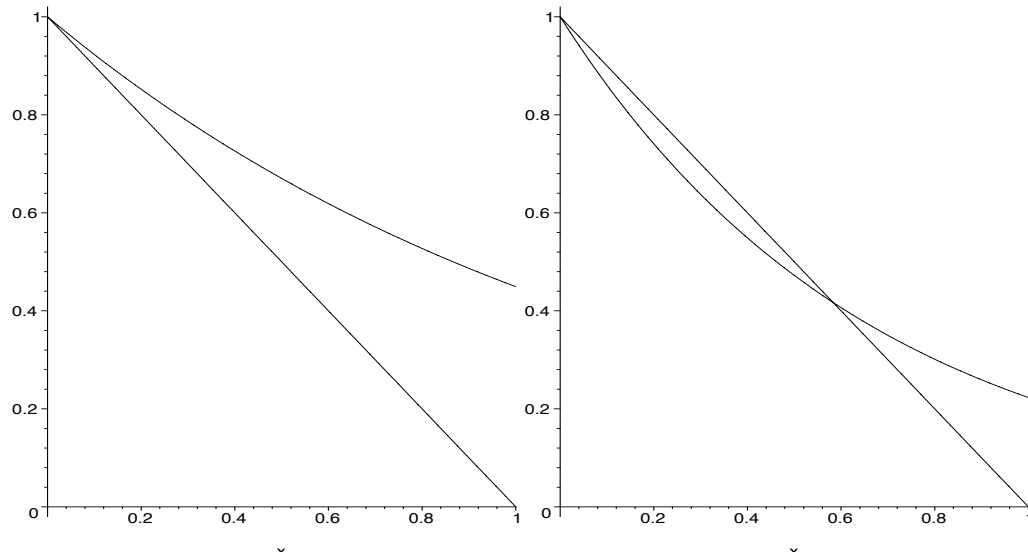
Tyvärr kan man inte lösa ut x ur denna ekvation. Det blir i stället en så kallad balansekvation som man får lösa numeriskt.

I figur 5 finns funktionerna $1 - x$ och $e^{-R_0 x}$ uppritade för två olika val av R_0 , nämligen $R_0 = 0,8$ och $R_0 = 1,5$. Lösningar till ekvationen $1 - x = e^{-R_0 x}$ ges alltså av punkter där kurvorna skär varandra. Vi ser att i den vänstra figuren ($R_0 = 0,8$) finns endast lösningen $x = 0$, men i den högra figuren finns två lösningar $x = 0$ och $x \approx 0,583$. Det går att visa matematiskt att om $R_0 \leq 1$ finns bara lösningen $x = 0$ men så fort $R_0 > 1$ finns även en annan strikt positiv lösning x^* . Hur stor x^* är beror på R_0 . För $R_0 = 1,5$ blir $x^* \approx 0,583$.

En balansekvation är en ekvation där den okända storheten finns på bägge sidor om likhetstecknet och man inte kan lösa ut den.

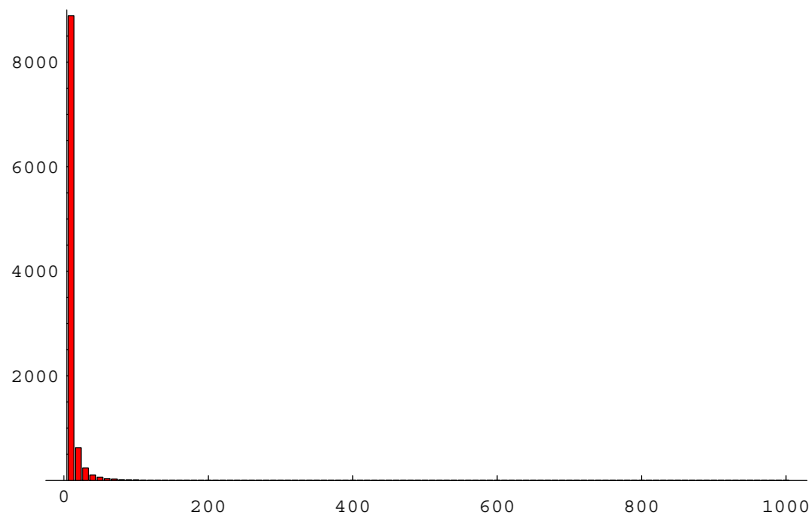
Vad är det då vi kommit fram till? Jo, för modellen ifråga kommer slutandelen smittade bli antingen ungefär 0, dvs epidemin tar aldrig fart, eller också tar epidemin fart och då smittas ungefär andelen x^* . Denna andel x^* är den positiva lösningen till ekvationen $1-x = e^{-R_0x}$.

Om $R_0 \leq 1$ finns ingen positiv lösning och epidemin kan inte ta fart. Jag har ju förstås inte visat resultatet rigoröst, något som dock kan göras men som kräver mer avancerad matematik.

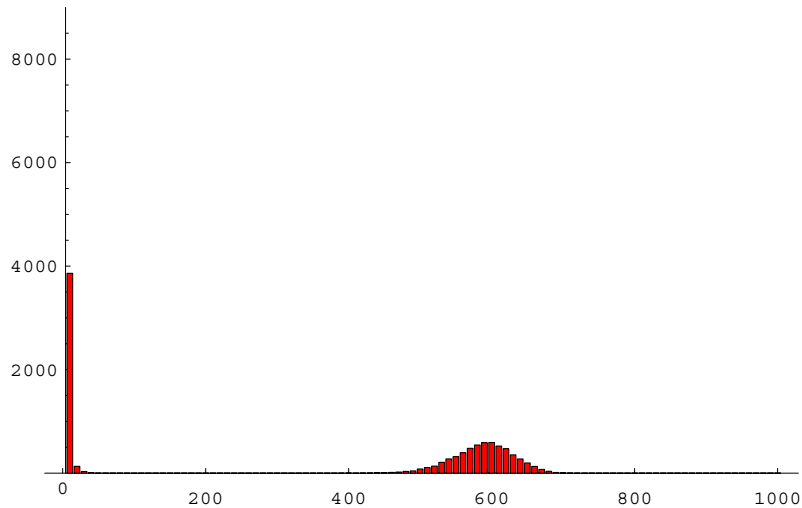


Figur 5. Funktionerna $f(x) = 1-x$ och $g(x) = e^{-R_0x}$ uppritade i samma graf. Vänstra figuren är för $R_0 = 0,8$ och högra för $R_0 = 1,5$. Som synes finns i vänstra figuren endast lösningen 0 till ekvationen $f(x) = g(x)$ medan det i högra figuren även finns en positiv lösning $x \approx 0,583$.

I figur 6 visas resultatet av 10 000 simuleringar av modellen i en population med 1000 individer för de två fallen $R_0 = 0,8$ och $R_0 = 1,5$. Histogrammen visar hur de 10 000 simuleringarnas slutstorlekar fördelade sig. Vi ser att när $R_0 = 0,8$ smittas endast ett fåtal i samtliga simuleringar. I fallet med $R_0 = 1,5$ ser vi däremot att det finns två kvalitativt olika möjligheter: antingen smittas endast ett fåtal, eller också smittas ungefär andelen $x^* \approx 0,583$. Bägge dessa iakttagelser stämmer med vad vi kommit fram till matematiskt. Den observanta anar för övrigt en liten normalfördelning kring $1000 \cdot x^*$. Även detta har bevisats strikt matematiskt, något jag inte närmare går in på.



Figur 6. Histogram över totala antalet smittade i 10 000 simuleringar av en population med 1000 individer. I en tänkt befolkning bestående av 1000 individer har man alltså 10 000 gånger simulerat smittomodellen startandes med en smittsam person, och i histogrammet framgår hur många som smittades totalt i de 10 000 simuleringarna. I vänstra figuren har simuleringarna gjorts med $R_0 = 0,8$ och i den högra, på nästa sida, med $R_0 = 1,5$. Observera att varje stapel har bredd 10, så tex andra stapeln innehåller antal simuleringar som resulterade i mellan 10 och 19 smittade totalt. Källa: Andersson & Britton.



Figur 6 forts

Statistiska slutsatser

Om vi observerar ett epidemiutbrott kan vi använda föregående resultat "baklänges", dvs om vi vet utbrottets storlek kan vi dra slutsatser om vilket värde reproductionstalet R_0 har. Om vi tex skulle se en epidemi i vilken 58% smittades kan vi dra slutsatsen att $R_0 \approx 1,5$. Mer allmänt får vi en skattning av R_0 genom att lösa ut R_0 i ekvationen som gavs för slutstorleken: $1 - x = e^{-R_0 x}$, där vi stoppar in observerad andel smittade som x . Detta ger $R_0 = -\ln(1 - x)/x$ (ln är den naturliga logaritmen), vilket för $x = 0,58$ ger $R_0 \approx 1,5$. I influensautbrottet som illustreras i figur 2 smittades 43 av totalt 129 individer, dvs 33%. Detta tyder på att $R_0 \approx -\ln(1 - 0,33)/0,33 \approx 1,23$.

Reproduktionstalet R_0 beror på hur smittsam sjukdomen ifråga är, men även på vilken population det gäller, tex tenderar tätbebyggda områden, speciellt i kombination med låg sanitär standard, ha högre R_0 än mer glest bebyggda regioner. Ett R_0 strax över 1 brukar vara vanligt för tex influensa. Barnsjukdomar som mässling och vattkoppor brukar däremot ha klart större R_0 mellan 10 och 20, dvs under en smittperiod har man mellan 10 och 20 smittsamma kontakter.

Vaccinering

Har man någon glädje av att veta vad R_0 är för en viss sjukdom och population? Svaret på frågan är ja. Om man vill undvika ett epidemiutbrott kan man göra detta genom vaccination, och R_0 avgör hur stor andel som måste vaccineras för att ett framtida utbrott ska undvikas. Antag tex att vi skulle vaccinera andelen v med ett vaccin som ger full immunitet mot en sjukdom med reproduktionstalet R_0 . (Kom ihåg att R_0 är genomsnittligt antal smittsamma kontakter en individ har under sin smittoperiod.) När andelen v av dessa är vaccinerade betyder det att genomsnittligt antal smittsamma kontakter med icke-vaccinerade är $R_0(1-v)$. Reproduktionstalet efter vaccinering betecknas R_v och har därför värdet $R_v = R_0(1-v)$.

Från tidigare vet vi att ett epidemiutbrott är möjligt endast om reproduktionstalet överstiger 1. Efter vaccination är detta möjligt endast om $R_0(1-v) > 1$. Vi skall således välja andelen vaccinerade v så att detta *inte* gäller eftersom en epidemi då inte kan ta fart, dvs v skall uppfylla $R_0(1-v) \leq 1$. Om vi löser ut v får vi villkoret

$$v \geq v_c = 1 - 1/R_0$$

Gränsvärdet $v_c = 1 - 1/R_0$ brukar kallas *den kritiska vaccinationsnivån*. I influensaexemplet från Minnesota i figur 2 hade vi i föregående avsnitt dragit slutsatsen att $R_0 \approx 1,23$. Således blir den kritiska vaccinationsnivån ungefär $v_c \approx 1 - 1/1,23 \approx 0,185$.

Detta betyder att om mer än 18,5% vaccineras på ålderdomshemmet i Minnesota kommer inget utbrott ske nästa gång influensaepidemin är på gång, förutsatt att vaccinet ger full immunitet. Det finns förstås en viss osäkerhet i vår uppfattning om R_0 ($\approx 1,23$) och i att vi gjort vissa förenklande antaganden. Hur som helst bör ålderdomshemmet

vara skyddat om minst ca 25 % skulle vaccineras. Man säger att hela ålderdomshemmet åtnjuter populationsimmunitet (jfr engelskans herd immunity). Anledningen till detta är att även icke-vaccinerade är skyddade så till vida att det inte kan bli något stort utbrott på ålderdomshemmet.

En av de viktigaste insikterna med matematiska modeller för epidemier är just relationen mellan R_0 och den kritiska vaccinationsnivån v_c . Speciellt viktig är insikten att man inte måste vaccinera alla individer, vilket i praktiken är omöjligt, för att skydda befolkningen mot framtida utbrott.

Modellering av endemiskt tillstånd

Betrakta återigen figur 3 som illustrerar mässling i England respektive Island. Figuren ger upphov till ett flertal intressanta observationer som i sin tur genererar frågor. Vad beror periodiciteten på? Varför är cykellängden 2 år i England men 4 år på Island? Hur kommer det sig att mässling var endemisk i England, dvs ej försvann helt, men kom i epidemier på Island? Hur många skulle man behöva vaccinera i England för att få mässlingen att försvinna?

Modeller för endemiskt tillstånd är matematisk krångligare än modeller för korta epidemiutbrott. För långa förlopp måste vi ta hänsyn till att individer föds och dör samt till immigration, medan man approximativt kan anta att population är konstant under en kort epidemi. Vi kommer inte här att lösa alla frågor med matematiskt resonemang. Frågan vi koncentrerar oss på är varför mässling var endemisk i England men ej på Island, samt hur många som behöver vaccineras – i England för att mässling ej längre skall vara endemisk och på Island för att undvika epidemiutbrott.

Först lite heuristiskt resonemang kring mässling och strukturen i figur 3. Som väl alla vet kan man bara ha mässling en gång i livet. För att mässling skall finnas endemisk i en population krävs således att nya individer tillkommer till populationen, något som sker genom födselar och immigration. En annan fråga är förstas varför denna periodicitet uppstår. Den förklaras av två faktorer. Om man tittar närmare på pucklarna i diagrammen ser man att de sätter fart på hösten, närmare bestämt i september vilket kanske inte syns i figurens

upplösning. Vad händer i september som kan påverka mässlingfrekvensen? Jo, det är då alla barn kommer tillbaka till skolan och daghemmet efter sommarlovet, och skolor och daghem fungerar mycket bra för att sprida mässlingen eftersom många barn möts i inte alltför stora lokaler. Men varför finns det då inte en puckel varje år? Förklaringen till detta är immunitet. Om det varit ett utbrott ett visst år smittas i stort sett alla barn i skolor och daghem som inte tidigare haft mässling. Året därpå är det således bara barn i första årskullen som inte fått mässlingen, och dessa är för få för att sätta igång ett större utbrott. Året därpå är det däremot både barn i ettan och tvåan som inte haft mässling, således fler, vilket tydligt möjliggör ett mässlingutbrott i England. På Island bor det färre varför det tycks som om man måste vänta hela fyra år innan det blivit tillräckligt många mottagliga i skolor och på daghem för att möjliggöra ett utbrott. Beskrivningen ovan är förstas en förenkling av verkligheten, men andemeningen är korrekt.

På Island, där smittan helt försvinner mellan utbrotten, orsakas ett nytt utbrott alltid av att en person (islänning eller av annan nationalitet) smittats i ett annat land och är smittsam när han eller hon kommer till Island. Det är inte så att detta sker endast vart fjärde år, utan det kommer förstas smittsamma till Island oftare. Dock måste tydligen fyra år ha gått sedan föregående epidemiutbrott för att tillräckligt många ska vara mottagliga för att sätta igång ett nytt utbrott och då endast i samband med skolstarten.

En endemisk population i jämvikt

Nu till varför en större population, så som England, hade mässling endemisk medan så inte var fallet på Island. För att besvara denna fråga gör vi förenklingen att vi inte har några säsongsvariationer av typen skolstart, och att vi har en population i jämvikt, både vad gäller populationens storlek, åldersstruktur och andel som smittats. Med jämvikt menas att dessa storheter inte varierar nämnvärt över tiden.

Liksom i föregående avsnitt definierar jag R_0 som genomsnittligt antal smittsamma kontakter under smittoperioden.

Om alla vore mottagliga är detta detsamma som det antal man i genomsnitt smittar. Jag ska nu härleda en relation mellan R_0 och andelen mottagliga i populationen.

Antag följande avrundade men någorlunda realistiska värden för en befolkning och mässling (innan vaccinering påbörjades):

- ▼ Genomsnittlig livslängd ≈ 75 år
- ▼ Genomsnittlig ålder då barn smittas av mässling ≈ 5 år
- ▼ Smittoperiodens längd ≈ 1 vecka $\approx 0,02$ år

En genomsnittlig människa lever 75 år och smittas av mässling när hon är 5 år gammal och har mässlingen 0,02 år efter vilket hon blir immun. Detta betyder att en individ spenderar andelen $5/75$ av sitt liv som mottaglig, andelen $0,02/75$ av sitt liv som smittsam och andelen $(75 - 5 - 0,02)/75$ av sitt liv som immun. Eftersom populationen är i jämvikt beskriver dessa siffror även andelen individer i populationen som befinner sig i respektive tillstånd:

- ▼ Andelen mottagliga i populationen $\approx 5/75 = 1/15$
- ▼ Andelen smittsamma i populationen $\approx 0,02/75 = 1/3750$
- ▼ Andelen immuna i populationen $\approx 1 - (1/15 + 1/3750) \approx 14/15$

Den översta relationen kan vi använda för att ta reda på vad reproduktionstalet R_0 är. En individ har i genomsnitt R_0 smittsamma kontakter under sin smittoperiod. Eftersom andelen mottagliga är ungefär $1/15$ betyder det att man i genomsnitt smittar $R_0 \cdot 1/15$ individer. Å andra sidan måste man ju i genomsnitt smitta exakt en ny människa för att populationen och mässlingens utbredning skall vara i jämvikt. Om varje individ i genomsnitt smittade t ex två individer skulle ju alltfler bli smittade vilket strider mot jämviktsantagandet att andelen som smittats var approximativt konstant. Om man å andra smittade mindre än en människa i genomsnitt skulle andelen smittade sjunka, även detta motsäger jämviktsantagandet. Således har vi härlett relationen

$$R_0 \cdot \frac{1}{15} \approx 1 \quad \text{dvs} \quad R_0 \approx 15 \approx 1/(\text{andel mottagliga}).$$

Denna siffra stämmer väl med vad som brukar anges för mässling, vanliga skattningar av R_0 för mässling ligger mellan 10 och 18. Vi kan även här använda vår kunskap om R_0 till att härleda ett uttryck för den kritiska vaccinationsnivån v_c . På samma sätt som tidigare gäller nämligen att

$$v_c = 1 - \frac{1}{R_0} \approx 1 - \text{andel mottagliga} \approx 1 - \frac{1}{15} = 14/15 \approx 93\%.$$

Från detta drar vi alltså slutsatsen att man bör vaccinera minst 93% för att omöjliggöra framtida utbrott av mässling. Om detta görs kontinuerligt, dvs även med nyfödda och immigrant-er kommer inte heller tillfälliga besökare att sätta igång några större epidemier. Denna nivå har uppnåtts med råge av flertalet länder i väst, och som en konsekvens finns heller inte mässling endemiskt. Endast enstaka mindre utbrott förekommer.

Den andra jämviktsrelationen, dvs att andelen smittsamma är ungefär $1/3750$, kan användas till att förstå varför sjukdomen blir endemisk endast i större länder. Om populationen består av n människor blir *antalet* smittsamma personer i genomsnitt $n/3750$, dvs

$$\text{genomsnittligt antal smittsamma} \approx \text{populationsstorleken}/3750.$$

För ett land som Island, som omkring 1960 hade ca 170 000 invånare, blir det i genomsnitt ca 45 smittsamma individer medan motsvarande siffra för England, med ca 50 miljoner invånare, blir ca 13 300 smittsamma individer. Det verkliga antalet smittsamma ligger förstås och pendlar kring detta genomsnitt. På Island pendlar alltså antalet smittsamma kring 45. Man inser då att det av en ren slump någon gång blir 0 som är smittsamma. När detta inträffat är landet fritt från smitta och det kommer att förbli 0 smittsamma, tills någon för in smittan från annat land. Detta förklarar alltså varför mässling inte är endemisk på Island. Antalet smittade i England ligger däremot och pendlar kring 13 300 smittsamma. Ett så stort värde kommer knappast ner till 0 av slumpen, vilket medför att sjukdomen blir endemisk.

Slutord

Vi har nu sett att man med matematik kan härleda viktiga slutsatser för smittsamma sjukdomars utbredning. Huvudresultaten har varit att härleda hur många som behöver vaccineras för att framtida epidemiutbrott ska undvikas baserat på hur många som tidigare smittades, och för det endemiska fallet hur stor vaccinationsgraden måste vara för att få bort sjukdomen baserat på den endemiska nivån. Jag har förenklat verkligheten en hel del så fäst inte alltför stor vikt vid de numeriska värdena. I litteraturen finns mer realistiska, och därmed mer komplicerade, matematiska modeller för vilka motsvarande resultat är mer tillförlitliga. Man kan dock med matematiska modeller aldrig nå ända fram till sanningen. Ett visst mått av osäkerhet kommer alltid att finnas.

Övningar

1. Försök att räkna ut sannolikheten för de olika möjliga slutstorlekarna om det finns fyra mottagliga och en smittsam individ från början.

2. Slutandelen smittade x för epidemimodellen i en stor population löste ekvationen $1 - x = e^{-R_0x}$. Givet R_0 kan man bara räkna ut lösningen numeriskt genom att saxa sig in. Starta tex med $x = 0,5$. Räkna ut vänsterled ($1 - 0,5 = 0,5$) och högerled ($e^{-R_0 \cdot 0,5}$). Om vänsterledet är större än högerledet gör man om samma sak, men med ett nytt x som är föregående plus halva föregående förändring (tag förändring $0,25$ första gången). Om vänsterledet är mindre än högerledet tar man i stället föregående x minus halva föregående förändring. Gör detta för fallet $R_0 = 10$ för att på så sätt komma fram till hur många som ungefär kommer att smittas.

3. Relationen $1 - x \approx e^{-R_0 x}$, som gäller för flera modeller, har en naturlig tolkning. Varje smittad individ ger (i genomsnitt) ifrån sig "smittrycket" R_0/n . Det är totalt nx smittade, så det totala smittrycket S blir således $S = nx \times R_0/n = R_0 x$. Chansen att inte smittas kan således skrivas som e^{-S} där S är smittrycket för modellen ifråga. En sådan modell föreligger när individer, utöver att smittas av en annan person, även kan bli smittade av annan orsak, t ex genom dåligt vatten eller via mat. Antag att individer smittas på detta sätt med sannolikheten e^{-a} , så "trycket" från detta sätt är a . Hur blir motsvarande jämviktsrelation för andelen smittade x i detta fall, dvs när man kan smittas av någon (sannolikheten R_0/n för given person) eller på detta andra sätt?

4. En mer realistisk modell för endemiska sjukdomar, som t ex mässling, tar hänsyn till att sjukdomen har en sk latensperiod. Man sägs vara latent om man är smittad men ännu inte blivit smittsam. För mässling är latensperioden i genomsnitt ca en vecka lång. Försök generalisera jämviktstillståndet för mässling med hänsyn taget till detta, dvs där individer är antingen mottagliga, latent, smittsamma eller immuna. En sjukdom försvinner från en population först när det varken finns smittsamma eller latent individer. Hur stor andel av befolkningen befinner sig i något av dessa tillstånd? Hur många kommer i genomsnitt att befinna sig i något av dessa tillstånd i England respektive Island? Jämför med modellen utan latensperiod. En lärdom från detta är att latensperiodens längd är lika viktig som smittperiodens längd för att avgöra om sjukdomen kommer att bli endemisk eller ej.

Litteratur

Några referenser på svenska som behandlar liknande frågeställningar finns inte. Vad gäller litteratur på engelska finns det däremot mycket att välja bland.

En referens som håller en enkel matematisk nivå och som fokuserar på sjukdomar:

Giesecke, J. (2002). *Modern infectious disease epidemiology, 2nd Ed.* London: Arnold.

En viktig referens som tar upp en mängd intressanta fenomen från verkligheten och som studerar dessa med hjälp av lite mer avancerade matematiska modeller och där man finner bla förteckningar över R_0 och v_c för ett flertal sjukdomar:

Anderson, R. M. & May, R. M. (1991). *Infectious diseases of humans: dynamics and control.* Oxford: Oxford University Press .

En mer statistiskt inriktad referens:

Becker, N. G. (1989). *Analysis of infectious disease data.* London: Chapman Hall.

En nyare bok som är mer avancerad och som förutsätter att man kan sannolikhetsteori och statistisk teori:

Andersson, H. & Britton, T. (2000). *Stochastic epidemic models and their statistical analysis.* New York: Springer.

