

# Smittsamma sjukdomars matematik

**Tom Britton**

---

Matematiska institutionen  
Stockholms universitet  
tom.britton@math.su.se

Jag ska nu beskriva dels lite om matematiska modeller för smittsamma sjukdomar men också hur dessa modeller används av hälsomyndigheter och beslutsfattare i konkreta situationer. Problemområdet har varit extra aktuellt senaste året i samband med den nya influensan (H1N1) pandemin. Som forskare i området känner jag mig dock inte orolig för att det här ska vara sista gången som liknande modeller kommer till användning. De senaste 20 åren har ett flertal sjukdomar varit lika aktuella: HIV, mul- och klövsjuka, hotet från bio-terrorism (t.ex. smittkoppor) och SARS för att nämna några.

Först beskriver jag ett par enkla modeller för smittspridning och därefter nämner jag olika generaliseringar som gör modellerna mer verklighetstroga.

## ***Två enkla modeller för smittspridning***

Vi vill konstruera en modell för sjukdomar som sprids från människa till människa via någon form av kontakt – inte sjukdomar som sprids indirekt t.ex. via mat eller vatten.

Innan man smittats av sjukdomen ifråga är man *mottaglig*. En mottaglig person kan förr eller senare smittas, och om detta händer blir han/hon *smittsam*. Som smittsam kan man smitta andra individer, och efter att ha varit smittsam en viss tid *tillfrisknar* man och blir immun, åtminstone ett tag. Många sjukdomar har förstås också viss dödlighet, jag kommer här bortse från detta – det matematiskt viktiga är att man upphör att sprida sjukdomen vidare. På engelska brukar de tre kursiverade tillstånden ovan betecknas susceptible-infectious-recovered, varför motsvarande klass av epidemimodeller brukar kallas SIR-modeller.

Nästa fråga vi måste ställa oss är huruvida vi vill modellera smittspridning på kort eller lång sikt, alternativt om vi intresserar oss för en ej ännu, eller nyligen, anländ sjukdom, eller om vi vill modellera en sjukdom som funnits länge i befolkningen, dvs som är *endemisk* (t.ex. vattkoppor). Det förra kallar vi för en *epidemimodell* och det senare för en *endemisk modell*. Epidemimodeller är enklare eftersom vi där har ett kort tidsperspektiv och därför kan anta att vi har en fix och given population – vi bortser alltså från födslar, dödsfall, immigration och emigration.

## En enkel epidemimodell

I den enklaste deterministiska epidemimodellen antar vi en fix och given population i vilken alla individer träffar alla lika ofta. Risken för en mottaglig individ att smittas är därför proportionell mot andelen smittsamma i population, och varje smittsam individ tillfrisknar och blir immun med konstant intensitet. Vi får således följande differentialekvationer för andelen mottagliga, smittsamma och tillfrisknade ( $s(t), i(t), r(t)$ ):

$$\begin{aligned} s'(t) &= -\lambda s(t)i(t), \\ i'(t) &= \lambda s(t)i(t) - \gamma i(t), \\ r'(t) &= \gamma i(t), \end{aligned} \tag{1}$$

med startvärden  $(s(0), i(0), r(0)) = (s_0, i_0, r_0)$ , där vi antar att  $i_0 > 0$ ; annars är alla derivater 0. Parametern  $\lambda$  är med andra ord intensiteten (i given tidsintensitet) med vilken smittsamma individer träffar andra individer och  $\gamma$  är intensiteten med vilket individer tillfrisknar och blir immuna (i samma tidsenhet);  $1/\gamma$  är således genomsnittlig tid man är smittsam. Eftersom alla individer är antingen mottagliga, smittsamma eller tillfrisknade så gäller för alla  $t$  att  $s(t) + i(t) + r(t) = 1$ ; differentialekvationssystemet är endast två-dimensionellt. Om man betraktar differentialekvationerna ser man att andelen mottagliga  $s(t)$  avtar hela tiden. Detta är inte förvånande eftersom vi har en fix population och när folk smittas minskar antalet mottagliga. Vad gäller andelen smittsamma  $i(t)$  kan differentialekvationen skrivas  $i'(t) = i(t)(\lambda s(t) - \gamma)$ . Det betyder att derivatan är proportionell mot andelen smittsamma medan andra faktorn avtar eftersom  $s(t)$  avtar. Tolkningen, och orsaken, till detta är att i takt med att andelen mottagliga avtar så "ödslas" allt fler potentiellt smittsamma kontakter på redan smittade personer – endast kontakter med mottagliga resulterar i att nya individer smittas. Tids nog blir andra faktorn negativ och andelen smittsamma börjar då avta.

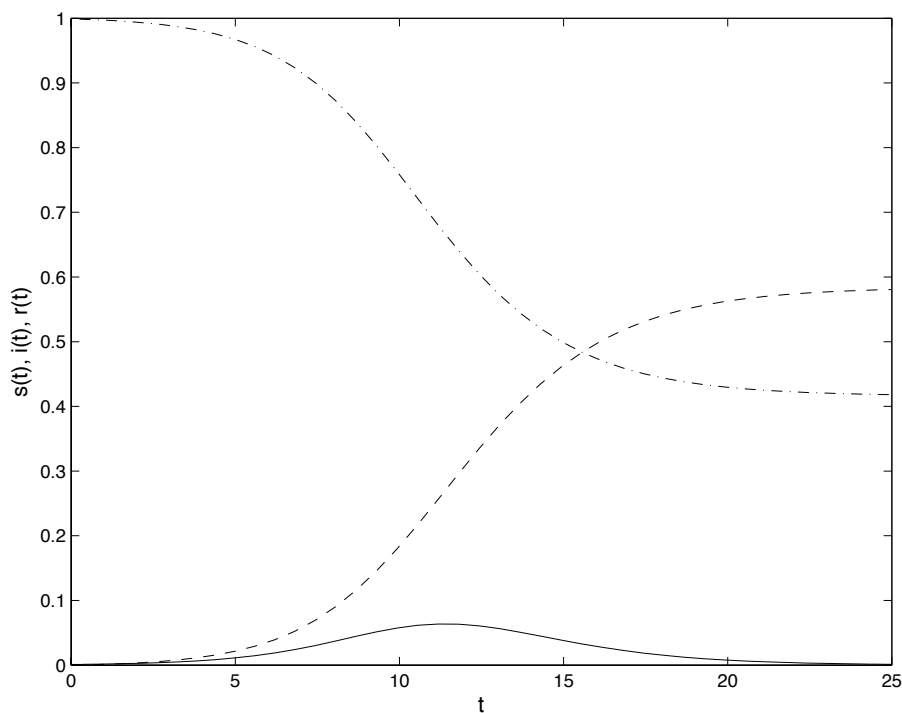
Differentialekvationerna är icke-linjära. Likväl kan man förvissa sig om att lösningen konvergerar mot  $(s_\infty, 0, r_\infty) = (1 - r_\infty, 0, r_\infty)$  då  $t \rightarrow \infty$ , där  $r_\infty$  löser ekvationen

$$\frac{1 - r}{s_0} = e^{-\frac{\lambda}{\gamma}(r - r_0)}. \tag{2}$$

Denna ekvation kan lätt tolkas: andelen av de initialt mottagliga som *inte* smittas under epidemin (vänster sida) måste klara smittrycket från dem som är smittsamma under epidemin (höger sida).

Om  $i_0 > 0$  (som vi antagit) så finns en entydig lösning  $r_\infty > r_0$  till denna ekvation. Ett numeriskt exempel är  $\lambda = 1.5$ ,  $\gamma = 1$  (som anses stämma ganska väl med den nya influensan med veckor som tidsenhet) och  $(s_0, i_0, r_0) = (0.999, 0.001, 0)$ , dvs en mycket liten andel initialt smittade och resten mottagliga. Lösningen till (2) blir i detta fall  $r_\infty \approx 0.583$ . Det betyder alltså att ca 58% av befolkningen kommer att smittas av epidemin. Som parentes kan nämnas att man bedömer att väsentligt färre kommer att smittas av den pågående nya influensan av flera orsaker: dels antas en viss andel av befolkningen vara immuna, dels förväntas vi minska smittspridningen med olika åtgärder så som t.ex. bättre hygien, och slutligen kommer många

att vaccineras. I Figur 1 är differentialekvationssystemet löst numeriskt för dessa värden och vi ser hur andelen mottagliga, smittsamma och tillfrisknade varierar över tiden för att asymptotiskt nå slutstadiet  $(s_\infty, i_\infty, r_\infty) \approx (0.417, 0, 0.583)$



**Figur 1:** Numerisk lösning till ekvationssystem (1) för fallet  $\lambda = 1.5$ ,  $\gamma = 1$  och startvärde  $(s_0, i_0, r_0) = (0.999, 0.001, 0)$ .  $s(t)$  prickad,  $i(t)$  heldragen och  $r(t)$  streckad.

Av speciellt intresse är fallet när  $r_0 = 0$  (dvs inga initialt immuna) och en väldigt liten andel initialt smittade. Det visar sig då att lösningen satisfierar  $r_\infty \approx 0$  om  $R_0 := \lambda/\gamma \leq 1$ , medan lösningen  $r_\infty$  är nedåt begränsad (hur liten  $i_0$  än är) i fallet  $R_0 > 1$ . Således är fallet  $R_0 = 1$  är ett tröskelvärde.

Parametern  $R_0 = \lambda/\gamma$  kallas inom infektionsepidemiologin för det fundamentala reproduktionstalet (eng. basic reproduction number). Tolkningen av  $R_0$  är att det är lika med antalet individer en smittsam person smittar i en helt mottaglig population: man smittar andra med intensitet  $\lambda$  (när alla är mottagliga) och i genomsnitt är man smittsam  $1/\gamma$  tidsenheter. Kvoten  $R_0 = \lambda/\gamma$  avgör med andra ord hur många som kommer att smittas och denna parameter är oberoende av tidsskala. De enskilda parametrarnas storhet påverkar bara hastigheten med vilken sjukdomen sprids. Reproduktionstalet beror förstas på hur smittsam själva sjukdomen är (hur mycket virus en smittsam person utsöndrar) men också på vilken population som betraktas: ju mer och tätare vi umgås ju mer risk för smittspridning. I böcker om infektionssjukdomar finns reproduktionstal angivna för många sjukdomar i specificerade regioner och tidsperioder.

Eftersom  $R_0 = \lambda \frac{1}{\gamma}$  anger genomsnittligt antal personer som en smittsam person smittar (i en helt mottaglig befolkning) under sin smittoperiod är det inte så förvånande att  $R_0 = 1$  är ett tröskelvärde: om man i genomsnitt smittar mer än en individ (i början av epidemin) så kommer epidemin ta fart, medan om man i genomsnitt smittar mindre än en individ så finns inga möjligheter för ett större utbrott.

Som parentes kan nämnas att denna modell, eller snarlika varianter, även används för modeller av ryktesspridning och informationsförmedling. Smitta svarar då mot att sprida ryktet/informationen och tillfrisknande mot att man tids nog slutar att sprida ryktet/informationen. Samma modeller kan även användas för spridning av datorvirus – där har även terminologin från infektionssjukdomar tagits i bruk.

### En enkel modell för endemiskt tillstånd

Modellen ovan var för ett kort tidsförlopp; vi antog att inga nya (mottagliga) individer fördes in i populationen. Då är epidemin dömd att upphöra eftersom andelen mottagliga hela tiden minskar.

Om vi i stället vill modellera vad som händer under en längre tid så är denna approximation inte längre rimlig. Ett enkelt sätt att i en modell ”tillföra” mottagliga individer och låta individer dö är att anta att nya (mottagliga) individer föds med en konstant intensitet  $\mu$ , och att individer dör med samma intensitet  $\mu$  (vare sig de är mottagliga, smittsamma eller tillfrisknade). Ekvationssystemet ovan modifieras då till följande:

$$\begin{aligned} s'(t) &= \mu - \lambda s(t)i(t) - \mu s(t), \\ i'(t) &= \lambda s(t)i(t) - \gamma i(t) - \mu i(t), \\ r'(t) &= \gamma i(t) - \mu r(t), \end{aligned} \quad (3)$$

med startvärden  $(s(0), i(0), r(0)) = (s_0, i_0, r_0)$ , där vi liksom tidigare antar att  $i_0 > 0$ . Det fundamentala reproduktionstalet  $R_0$  ändras aningen i denna modell: man kan ju nu lämna smittstadiet av två orsaker: att man tillfrisknar eller att man råkar dö. Av denna anledning får man  $R_0 = \lambda/(\gamma + \mu)$ . Detta differential-ekvationssystem visar sig också ha olika karaktär beroende på om  $R_0$  är större än 1 eller inte: Om  $R_0 \leq 1$  så har ekvationerna endast ett (stabilt) jämviktsläge:  $(s(t), i(t), r(t)) = (1, 0, 0)$ , dvs att sjukdomen kommer försvinna ur befolkningen. Om däremot  $R_0 > 1$  så finns två jämviktslägen:  $(s(t), i(t), r(t)) = (1, 0, 0)$  som inte längre är stabilt, samt ett stabilt jämviktsläge

$$(s(t), i(t), r(t)) = (\hat{s}, \hat{i}, \hat{r}) = \left( \frac{1}{R_0}, \frac{R_0 - 1}{R_0 \delta}, 1 - \frac{1/\delta + R_0 - 1}{R_0 \delta} \right), \quad (4)$$

där  $\delta = \mu/(\gamma + \mu)$  är den (lilla) genomsnittliga andelen av livet som man är smittsam. Detta jämviktsläge kallas den *endemiska nivån*.

Som ett numeriskt exempel kan vi ta vattkoppor som är en endemisk sjukdom i Sverige (dvs det finns hela tiden smittbärare). För vattkoppor brukar  $R_0$  antas vara ungefär 15, och tiden man är smittsam är ca 1 vecka, så andelen av ens liv man är

smittsam (nästintill alla smittas!) är ca  $(1/52)/80 \approx 0.00024$  om vi lever 80 år i genomsnitt. Det kan även vara värt att nämna att andelen mottagliga, vilket sammanfaller med den relativa åldern som man smittas, ges av  $\hat{s} = 1/R_0 \approx 0.067$ . Så fort en endemisk sjukdom har stort  $R_0$  kommer alltså individer typiskt smittas i en tidig ålder. Detta förklarar benämningen "barnsjukdomar". Dessa förklaras alltså huvudsakligen av att de har stort  $R_0$  snarare än av barn skulle var extra mottagliga. Den s.k. endemiska nivån blir hur som helst i detta fall  $(\hat{s}, \hat{i}, \hat{r}) \approx (0.067, 0.00022, 0.933)$ . Vid första anblicken kan det vara förvånansvärt många tillfrisknade (=immuna), men, en typisk individ smittas vid ungefär 5 års ålder, är sedan smittsam en vecka och sedan tillfrisknad (och immun) resten av livet. I själva verket kan denna kunskap (om genomsnittlig ålder man smittas) användas för att skatta  $R_0$ . Om vi vet att individer smittas vid ca 5 års ålder och lever ca 80 år så är ju andelen mottagliga i befolkningen ca  $5/80$ , så  $5/80 \approx 1/R_0$ , dvs  $R_0 \approx 80/5 = 16$ .

## Prevention

Vi har nu visat hur man kan modellera ett epidemiutbrott respektive ett endemiskt tillstånd, förvisso under det förenklande antagandet att hela populationen är homogen och blandar sig likformigt. Dessa förenklingar har dock inte visat sig avvika från verkligheten alltför mycket för sjukdomar som just influensa, förkyllning och barnsjukdomar. För t.ex. sexuellt överförbara sjukdomar duger däremot inte dessa modeller som approximation av verkligheten (se avsnittet om nätverksmodeller nedan).

Något som också kan vara intressant att modellera (och som förstås infektions-epidemiologer är primärt intresserade av!) är hur man minskar smittspridningen och allra helst stoppar den. Grovt sett kan smittspridningen minskas på två sätt. Det ena sättet är att reducera  $R_0$  vilket kan göras på många olika sätt. T.ex. ombeds vi nu i nya influensans tider att tvätta händerna ofta, att nysa i armvecket m.m., och för allvarligare sjukdomar som SARS genom isolering. Det andra sättet är att minska andelen mottagliga, typiskt genom att vaccinera och på så sätt föra över individer från att vara mottagliga till att bli immuna (vilket innefattas i tillståndet *tillfrisknad*).

Om man ska försöka reducera  $R_0$  handlar det om att sträva mot att få ner det under 1. Vilka åtgärder som är mest (kostnads-)effektiva beror förstås på vilken sjukdom som betraktas. Om man ska vaccinera kan man använda lite mer matematik. Antag för enkelhets skull att vaccinet är perfekt, dvs att man blir helt immun, och låt  $v$  beteckna andelen som vaccineras innan epidemin anländer. Detta betyder att andelen immuna,  $r_0$ , ändras från att vara 0 till att bli  $v$ . Tröskelvärdet i Ekvation (2) ändras då från  $R_0 = 1$  till  $R_0(1-v) = 1$  (nu är ju endast andelen  $1-v$  mottagliga). För givet  $R_0$  betyder detta att om  $v \geq 1 - 1/R_0$  så är vi under tröskelvärdet och en försumbar andel kommer att smittas av epidemin. Gränsvärdet

$$v_c = 1 - 1/R_0$$

kallas för den kritiska vaccinationsgraden. För exemplet med influensa betyder detta att  $v_c \approx 1 - 1/1.5 \approx 0.33$ ; om man vaccinerar mer än 33% av befolkningen

med ett perfekt vaccin (innan sjukdomen har anlänt) så kommer *hela* befolkningen att vara skyddad, ett tillstånd som på engelska brukar kallas “herd immunity”. Således skyddar man fler än sig själv genom att vaccinera sig (nämligen dem man hade smittat om man inte vaccinerats och dem som dessa i sin tur hade smittat osv.).

För endemiska sjukdomar hjälper det inte att massvaccinera vid ett enstaka tillfälle. I stället måste man i detta fall kontinuerligt vaccinera en viss andel för att sjukdomen skall försvinna ur befolkningen. Om man vaccinerar andelen  $v$  kontinuerligt påverkar det (3) genom att första termen för  $s'(t)$  ändras till  $\mu(1 - v)$  och termen  $\mu v$  tillkommer i uttrycket för  $r'(t)$ . Det nya tröskelvärdet kan då visas bli  $R_0(1 - v)$  precis som i epidemimodellen, och om detta är mindre än 1 så finns bara ett (stabil) jämviktstillstånd, nämligen  $(\hat{s}, \hat{i}, \hat{r}) \approx (1 - v, 0, v)$ , dvs sjukdomen har försvunnit och populationsandelen mottagliga (=ej vaccinerade) är  $1 - v$  och resten är immuna (=vaccinerade). Den kritiska vaccinationsgraden blir som tidigare  $v_c = 1 - 1/R_0$ . Som exempel kan vi ta mässling som också brukar antas ha  $R_0 \approx 15$ . Den kritiska vaccinationsgraden är således  $v_c \approx 1 - 1/15 \approx 93\%$ . I Sverige har vi för tillfället en vaccinationsgrad på mer än 95%, och vaccinet är nästintill perfekt, varför vi heller inte har några mässlingsfall att prata om (utom i enstaka delpopulationer som av en eller annan anledning inte vaccinerar sig). Om man vaccinerar nyfödda kontinuerligt men inte kommer upp i den kritiska vaccinationsgraden  $v_c$  så kommer effekten bli att de som inte har vaccinerats typiskt kommer att smittas senare i livet. För vissa sjukdomar är symptomen väsentligt allvarligare när man är äldre, varför detta måste beaktas när nya vaccinationsprogram initieras.

## Stokastiska modeller

Jag har hittills inte nämnt stokastiska modeller (mitt eget forskningsområde) eftersom deterministiska modeller är enklare och många viktiga insikter kan hämtas från dessa. I vissa sammanhang är dock stokastiska modeller att föredra, några av dessa ska jag nu försöka beskriva kortfattat.

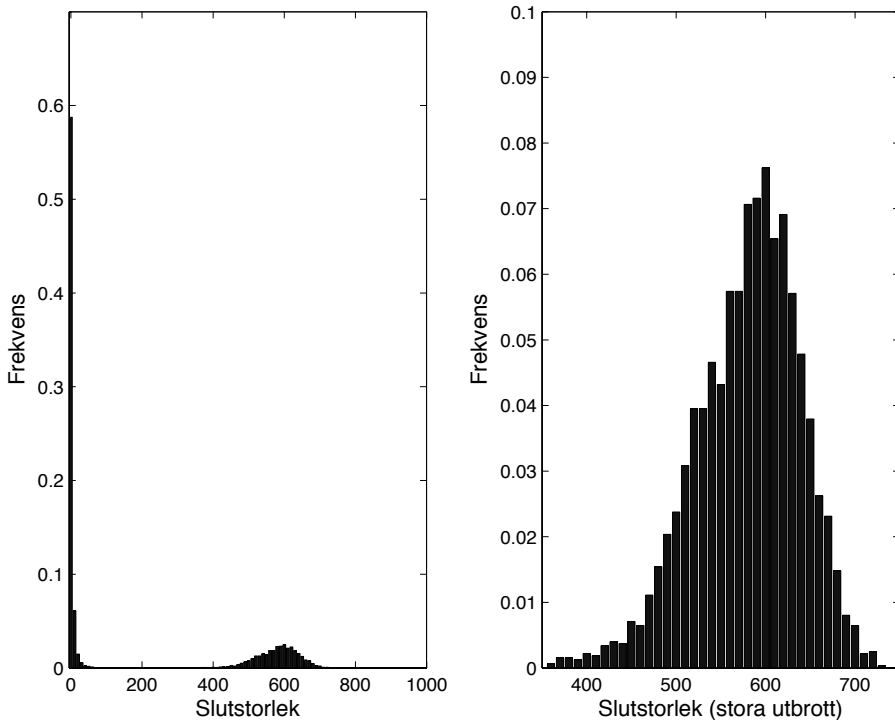
Låt mig först beskriva den allra enklaste stokastiska modellen och illustrera att den blir krånglig att räkna med även i små populationer. Förenklingen ligger i att vi antar att alla som smittas blir lika smittsamma och att man smittar varje annan individ oberoende av varandra och med samma sannolikhet – det finns förstås även modeller som släpper på dessa antaganden, men de blir än mer komplicerade. Antag följande modell som kallas Reed-Frost-modellen efter dem som kom på den. En individ som blir smittad smittar andra mottagliga individer, oberoende av andra, med sannolikhet  $p$  (samma  $p$  för alla individer). Efter detta tillfrisknar individen och blir immun för resten av studieperioden. De som smittas smittar i sin tur andra mottagliga med sannolikhet  $p$  osv. Antag att det inledningsvis finns 1 smittsam person (indexfallet) och  $n$  mottagliga.

Hur många av de initialt mottagliga kommer smittas innan epidemin dör ut? Eftersom det är en stokastisk modell kan vi inte veta exakt. Det enda vi säkert vet är att antalet initialt mottagliga som smittas under utbrottet,  $Z_n$ , ligger mellan 0 och  $n$ . Låt oss räkna ut sannolikheten för de olika möjliga utfallen när  $n$  är litet. Vi

börjar med det enklaste fallet  $n = 1$ . Vi har då inledningsvis 1 smittsam och 1 mottaglig. Sannolikheten att den mottaglige smittas blir ju då  $p$ , dvs  $P(Z_1 = 1) = p$ , och sannolikheten att han/hon inte smittas blir  $1 - p$ , dvs  $P(Z_1 = 0) = 1 - p$  (indexet "1" syftar på att  $n = 1$ ). För  $n = 2$  blir det aningen krångligare. För att ingen skall smittas får indexfallet inte smitta någon av de två mottagliga, så  $P(Z_2 = 0) = (1 - p)^2$ . För att exakt en skall smittas måste indexfallet smitta en av de två mottagliga (sannolikhet  $\binom{2}{1}p(1 - p)$ ) men dessutom får inte den då smittade smitta den kvarvarande. Sannolikheten att exakt en skall smittas blir således  $P(Z_2 = 1) = 2p(1 - p)^2$ . Sannolikheten att bägge smittas kan beräknas på två sätt: antingen som ett minus övriga sannolikheter, eller också genom att "räkna" ut sannolikheten. Det kan ske genom att indexfallet smittar en och den smittar den sista, eller att indexfallet smittar bägge. Den totala sannolikheten blir  $P(Z_2 = 2) = 2p^2(1 - p) + p^2$ . Redan när  $n = 2$  krävs således viss tankeförmåga. När vi låter  $n$  bli större blir beräkningarna snabbt oöverskådliga. Det går inte att få några enkla slutna uttryck för dessa sannolikheter. De kan visserligen beräknas numeriskt för givna värden på  $n$  och  $p$ , men dessa beräkningar är svåra att använda för att dra kvalitativa slutsatser. Detta motiverar behovet av approximationer när  $n$  är stort. I en stor population är det ofta mer naturligt att betrakta medelantalet  $np$  en person smittar som konstant snarare än smittosannolikheten  $p$  med en enskild individ. Man brukar därför låta  $p = R_0/n$ , medelantalet blir ju då  $np = R_0$ , dvs det reproduktionstal som tidigare definierades för deterministiska modeller. Vi ska nu kort beskriva hur approximationen ser ut. Vi använder enbart heuristiska resonemang även om motsvarande gränsvärdessatser finns.

Ett (potentiellt) epidemiutbrott inleds ofta med att ett fåtal individer smittats "utifrån" (i stycket ovan antog vi 1 indexfall). Det är ju då inte säkert att ett utbrott tar fart. Inledningsvis spelar därför stokastiken en viktig roll. Om populationen  $n$  i betraktelse är stor (och individer blandar sig helt slumpmässigt) är det mycket osannolikt att två smittokontakter skulle ske med samma individ. Från detta följer att smittsamma individer (i början) smittar andra individer nästintill oberoende av varandra. Epidemin kan därför inledningsvis approximeras av en s.k. förgreningssprocess. Med hjälp av teori för förgreningssprocesser kan man därför beräkna sannolikheten för att en epidemi tar fart, respektive sannolikheten för att endast få smittas. Man kan då även införa lite mer realism i modellen genom att låta individer vara olika smittsamma (t.ex. genom att vara olika sociala, ha olika infektivitet och/eller olika ha långa långa smittperioder). Detta kommer i hög grad påverka sannolikheten för ett stort utbrott, medan slutstorleken vid ett stort utbrott primärt beror av motsvarande genomsnittsvärden. I fallet att det blir ett stort utbrott kan man även visa en central gränsvärdessats som säger att slutandelen smittade blir normalfördelad med väntevärde  $r_\infty$  (lösningen till Ekvation 2 ovan) och med en explicit standardavvikelse av storleksordning  $1/\sqrt{n}$ . Även antalet smittade blir normalfördelat om det blir ett stort utbrott, med väntevärde  $nr_\infty$  och en standardavvikelse av storleksordning  $\sqrt{n}$ . Detta resultat illustreras i Figur 2 nedan. Där har 10 000 epidemier simulerats, alla i en population bestående av 1000 individer varav initialt 1 varit smittsam och övriga 999 mottagliga. Simuleringen är gjord för  $R_0 = \lambda/\gamma = 1.5$ . Om man löser Ekvation (2) för detta  $R_0$  (och med  $s_0 = 0.999$  och  $r_0 = 0$ , dvs inga initialt immuna och andelen 0.999 initialt mottagliga) får man lösningen  $r_\infty \approx 0.583$ . Det betyder alltså att slutantalet smittade bör vara ungefär

$nr_\infty \approx 583$  vid ett stort utbrott. Detta stämmer ganska väl med simuleringarna! Histogrammet till vänster anger frekvensfördelningen av slutantalet smittade i epidemin. Som synes består denna frekvensfördelning tydligt av två olika delar: de små utbrotten och de stora utbrotten. I figuren till höger tar vi en närmare titt på frekvensfördelningen för de stora utbrotten. Som synes är denna fördelning inte helt olik normalfördelningen – hade vi simulerat en större befolkning hade likheten varit än större. Väntevärdet i normalfördelningen tycks mycket riktigt ligga ungefär vid 583 – i alla fall strax under 600. Resultatet att slutandelen smittade vid en större epidemi är normalfördelat, är även viktigt när man ska skatta parameterar (t.ex.  $R_0$  och den kritiska vaccinationsgraden  $v_c$ ) från tidigare utbrott. Man får då inte bara punktskattningar  $\hat{R}_0$  och  $\hat{v}_c$ , utan även konfidensintervall på skattningarna.



**Figur 2:** Histogram över slutstorleken från 10000 simuleringar av en stokastisk smittomodell med  $R_0 = 1.5$ . Till vänster redovisas samtliga simuleringar och till höger de simuleringar som resulterade i stora utbrott.

En annan situation då stokastiska modeller är viktiga är om populationen inte blandar sig homogent, utan där det i stället finns små “kluster” inom vilka smittspridningen är väsentligt högre. Då är deterministiska modeller med “massverkan” olämpliga approximationer av verkligheten. Huvudexemplet på sådana kluster för bl.a. influensa och förkylning är hushåll/familjer. Det är ju mycket större risk att jag blir smittad av mina barn än att jag smittas av någon på gatan. I sådana fall bör man i modellen ta hänsyn till stokastiken inom hushållen medan man kan modellera smittspridningen mellan hushåll med en deterministisk modell. En annan



klass av sjukdomar där sociala strukturer spelar stor roll är sexuellt överförbara sjukdomar. Vill man modellera sådana bör man studera smittspridning på nätverk, där kanterna i nätverket representerar sexuella kontakter mellan människor (se nedan).

Ett sista exempel jag vill nämna där stokastiken är viktig är vid utdöende endemiska sjukdomar. Vi såg ju i exemplet med mässling att den endemiska nivån var  $(\hat{s}, \hat{i}, \hat{r}) \approx (0.067, 0.00022, 0.933)$ . Det betyder alltså t.ex. att andelen smittsamma i befolkningen en given tidpunkt bör vara ungefär  $\hat{i} = 0.00022$ . I ett land som Island med ca  $n = 250\,000$  individer (före vaccinet infördes på 60-talet) betyder det att *antalen* i endemiska nivån är ca  $(n\hat{s}, n\hat{i}, n\hat{r}) \approx (16\,700, 50, 233\,250)$ . Antalet smittsamma en given tidpunkt (när populationen och sjukdomen är i jämvikt) är således endast 50 personer! Självklart varierar detta över tiden, så det är inte förvånande att antalet smittsamma ganska snart råkar gå ner i 0 (bl.a. finns ju säsongsvariation som jag inte behandlat), och när detta väl inträffat så är sjukdomen utplånad. Den kloke inser dock att Island knappast var helt isolerat ens före 60-talet, så mässlingen borde snabbt ha förts tillbaka in i landet igen. Detta är bara delvis rätt tänkt. En ny epidemi kan nämligen inte ta fart direkt efter att sjukdomen dött ut – det finns då för få mottagliga. Efter en tid har dock tillräckligt många (mottagliga) barn fötts för att ett nytt epidemiutbrott kan ske. Jag har självklart förenklat verkligheten, men likväl stämmer det om man tittar på mässlingsfall på Island under perioden före vaccination infördes: att stora utbrott skedde med exakt 4 års intervall (regelbundenheten beror på skolstarten efter sommaren) och endast enstaka importerade fall däremellan! I England, med en väsentligt större befolkning, blir den endemiska nivån på antalet smittsamma  $n\hat{i} \approx 11\,000$ . Detta antal kommer förstas också fluktuera, men det är inte lika troligt att det slår i nollan inom överskådlig tid. Även detta bekräftas av data som visar att mässlingen hela tiden var endemisk i England under den aktuella perioden. Att försöka förstå vilken den ”kritiska populationsstorleken” är (som separerar dessa två olika scenarier), och hur den beror av olika egenskaper i populationen, hos sjukdomen och även vaccinet i förekommande fall, är ett forskningsområde jag själv är intresserad av.

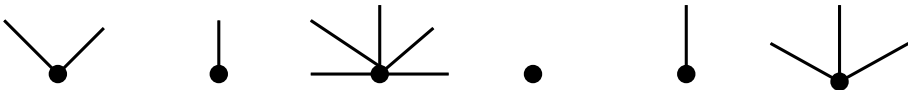
## ***Smittspridning på sociala nätverk***

Vi ska nu avslutningsvis bland olika modeller ta en titt på modeller för smittspridning där sociala strukturer anses alltför viktiga för att man ska kunna anta att alla smittar alla med ungefär samma sannolikhet (likformig smittspridning). Självklart vet man att likformig smittspridning aldrig är sann, men likväl visar erfarenheten att dessa modeller fungerar bra för mycket smittsamma sjukdomar så som t.ex. våra vanliga barnsjukdomar och förkylning, där ”smittokontakten” kan innebära att man vistas i samma rum/buss eller liknande. Andra sjukdomar kräver närmare kontakt för att smittoöverföring över huvud taget ska vara möjlig. En sådan klass av sjukdomar är sexuellt överförbara sjukdomar – där är det uppenbart att ingen människa har samma smittsannolikhet för alla andra människor. Det kan då vara lämpligt att först försöka modellera hur det sociala nätverket ser ut (detta ”vet”

man sällan) och därefter studera hur en smittsam sjukdom kan tänkas spridas "på" nätverket. Jag ska nu kort redogöra för en enkel sådan modell som likväl ger intressanta slutsatser.

Antag (för enkelhets skull) att vi har en enkönad befolkning och att somliga par av individer är sammanbundna med kanter (svarande t.ex. mot att de haft en sexuell relation det senaste året). Antalet andra individer en person har en kant till kallas personens *grad*. Alla individer har förstås inte samma grad. Hur graderna varierar i befolkningen beskrivs av den s.k. gradfördelningen. Om vi låter slumpvariabeln  $D$  vara graden hos en slumpvis vald individ så beskrivs gradfördelningen av  $p_k = P(D = k)$  för alla icke-negativa  $k$ . Låt  $\mu = E(D) = \sum_k k p_k$  beteckna väntevärdet (medelvärdet) i denna fördelning och  $\sigma^2 = V(D) = \sum_k (k - \mu)^2 p_k$  fördelningens varians. Utöver denna gradfördelning antar vi enklast möjliga slumpmässiga nätverk/graf: dvs individer väljer sina kant-partners helt slumpmässigt. På detta nätverk antar vi vidare att en person smittas utifrån (indexfallet) och att smittan sprider sig i nätverket genom att smittade personer smittar var och en av sina ännu ej smittade grannar med sannolikhet  $p$  och att detta sker oberoende.

En viktig fråga är förstås vad  $R_0$  är för denna modell, dvs. hur många smittar en typisk smittad individ i inledningsfasen av epidemin? En första naturlig gissning är att  $R_0 = \mu p$ , dvs att en typisk smittad person har grad  $\mu$  och att man smittar andelen  $p$  av dessa. Detta resonemang innehåller två fel. Först och främst blev man ju smittad av en sina grannar vilket motiverar att vi måste reducera genomsnittliga graden med 1. Det andra fallet ligger i att den genomsnittliga graden bland dem som smittas i inledningsfasen *inte* är  $\mu$ . Titta på det lilla exemplet nedan (Figur 3).



**Figur 3:** Illustration av 6 individer med olika antal grannar, dvs olika gradtal.

Där finns 6 individer med olika grader (antal kanter). Vem skulle du tro har störst risk för att bli smittad? Den som inte har någon kant kommer ju t.ex. helt säkert inte att bli smittad. Man kan i själva verket visa att sannolikheten att smittas i inledningsfasen är proportionell mot antalet grannar man har. Vi har således en annan gradfördelning  $\tilde{D}$  bland de initialt smittade än i befolkningen i stort (i vardagligt tal talar man om riskgrupper). Att sannolikheten att smittas är proportionell mot antalet grannar man har ger en entydig relation mellan fördelningarna  $\tilde{D}$  och  $D$ :  $P(\tilde{D} = k) = kP(D = k)/\mu$ . Man kan lätt visa att denna fördelning har väntevärde som kan uttryckas i väntevärdet  $\mu$  och variansen  $\sigma^2$  för populationens gradfördelning:  $E(\tilde{D}) = (\sigma^2 + \mu^2)/\mu$ . Dock kan man ju inte smitta den som man själv smittades av, så  $R_0 = p(E(\tilde{D}) - 1)$ , dvs andelen  $p$  av detta väntevärde subtraherat med 1. Om man förenklar formlerna lite får vi därför

$$R_0 = p \left( \mu + \frac{\sigma^2 - \mu}{\mu} \right) = p \left( E(D) + \frac{V(D) - E(D)}{E(D)} \right). \quad (5)$$

Vad kan vi då lära oss av denna relation? Den första termen  $pE(D)$  var vår första gissning av  $R_0$ , dvs populationens genomsnittliga grad multiplicerat med sannolikheten att smitta en enskild granne. Den andra termen innehåller däremot variansen i populationens gradfördelning. Om vi har stor varians kommer denna term dominera den första. Vi kan alltså ha  $R_0 > 1$  (som möjliggör ett större epidemiutbrott liksom att sjukdomen kan bli endemisk) även om smittsannolikheten  $p$  är liten OCH populationen har en låg genomsnittlig gradfördelning, så länge populationens gradfördelning har stor varians (spridning). Förklaringen till detta är att om gradfördelningen har litet väntevärde men stor varians, vilket betyder att det finns ett fåtal individer med mycket högt gradtal, så kommer typisk dessa individer att smittas i inledningsfasen. Och, trots att smittsannolikheten är låg så kommer dessa personer i sin tur smitta mer än en person i genomsnitt eftersom de har en så hög grad.

Låt oss avsluta med ett numeriskt exempel för en sexuellt överförbar sjukdom. Vi kan t.ex. anta att  $p = 0.3$ , dvs sannolikheten att smitta en partner är 0.3, att man är smittsam under ett år, och att individer i populationen i genomsnitt har  $\mu = 2$  partners under ett år. Kan vi ha ett utbrott för en sådan sjukdom? Vid första anblicken verkar det otroligt: om man har två partners och smittar en partner med sannolikhet 0.3 kan väl inte individer i början av utbrottet smitta mer än en person så att en epidemi tar fart? Svaret är emellertid att det visst kan hända, om populationens partnerfördelning är tillräckligt skev så att dess varians blir stor. Antag t.ex. att 20% av befolkningen har ingen partner under ett år, 40% har en partner, 30% har två partners medan de kvarvarande 10% har 10 partners (vårt exempel är inte realistiskt, men att några få har många partners stämmer dock med verkligheten). Genomsnittligt antal partners blir då just  $\mu = 2$ . Man kan beräkna variansen som blir  $\sigma^2 = 7.6$ . Med hjälp av Ekvation (5) får vi då att  $R_0 = 4.8/3 = 1.6 > 1$  vilket alltså möjliggör risken för ett stort utbrott och att sjukdomen blir endemisk.

För denna och liknande modeller kan man alltså visa att det räcker att gradfördelningen har stor varians för att risken för stora utbrott skall finnas. Det är emellertid inte så att det nödvändigtvis blir väldigt stora utbrott. I själva verket finns en tendens att andelen som smittas blir ganska liten om gradfördelningen är skev, de som smittas är primärt de med många partners (stor grad). Detta gäller i ännu högre grad om individer med hög grad tenderar att välja partners som också har hög grad. Då kommer sjukdomen i än högre grad koncentreras bland dessa "riskgrupper".

## **Verkligheten**

Jag har ovan beskrivit några enkla modeller och gett ett par "leksaksexempel". Man kan då fråga sig om metoderna inte är mer sofistikerade när mycket står på spel, som t.ex. nu för den pågående nya influensan. Svaret är, som så ofta, att de enkla modellerna ger de viktiga insikterna, men sedan gör man förstås mer verklighetstroga modeller när de ska användas i specifika sammanhang för att få mer precisa kvantitativa slutsatser.

Det finns många aspekter som brukar tillföras modeller som t.ex. ska förutsäga hur många som kommer smittas av den nya influensan, titta t.ex. i någon av tidskrifterna *Nature* eller *Science*. Några exempel på sådana är: effekter inte bara av hushåll, men också skolor/dagis och ibland även arbetsplatser, samt information om rese-mönster. Att ha en spatial struktur är också viktigt (en effekt blir att tillväxttakten i epidemin snarare blir linjär än exponentiell) och att man också tar hänsyn till individers olika mottaglighet och sociala struktur över åldern: barn tenderar att vara mer mottagliga och ha tätare kontakter än vuxna, samt att merparten individer umgås i högre grad med jämnåriga. Säsongsvariationer är en annan viktig aspekt att ta hänsyn till om man vill komma verkligheten nära. I modellen för endemiska sjukdomar antog vi en befolkning i jämvikt. I verkligheten är förstas inte åldersstrukturen helt jämn i populationen och populationer tenderar att växa, vilket är något som bör tillföras en modell för ökad realism. I modeller för smittspridning bland växter är den spatiala strukturen mycket viktigare än den är för människor. Eftersom växter inte rör sig är det förstas absolut störst chans att sprida smittan till sina närmaste grannar.

För mer komplicerade modeller applicerade på specifika länder (och deras befolkningsstruktur) är det förstas svårt att dra några analytiska slutsatser. Vad man i stället gör är stora simuleringsstudier, där man även varierar parametrar, för att se vad som kan tänkas hända. Nästintill alltid studerar man också vilka effekter olika preventiva åtgärder (vaccinering, men också stängning av drabbade skolor m.m.) kan tänkas ha. I komplicerade modeller med många parametrar har parametrarnas värden förstas osäkerhet. Det är inte alltid lätt att avgöra vilka parametrar som har störst inflytande och vilka de rimliga värdena för dessa parametrar är i en given befolkning. Genom att göra noggrann s.k. sensitivitetsanalys i kombination med god matematisk förstaselse för modellerna kan simuleringar dock tillföra viktiga insikter om spridningen av sjukdomar och vilka de effektiva preventiva åtgärderna är.

Jag var i höstas på ett möte om H1N1-influensan på europeiska smittskyddsinstitutet ECDC, som råkar ligga i Stockholm, och tog del av flera liknande modeller och hörde vad infektionsepidemiologer hade att säga. En intressant sådan studie visade att om man kunde implementera vaccineringen i USA (där endast en mindre andel kommer att vaccineras) innan epidemin tar riktig fart så kommer detta att reducera antalet som smittas (och antalet som dör) med 2/3 jämfört med om epidemin redan tagit fart innan vaccineringen kommit igång. Slutsatsen var att det är viktigt att skynda på framställandet av vaccinet och att snabbt distribuera det, men även att åtgärder syftandes till att *fördröja* smittspridningen är viktiga (dylka åtgärder är mycket mindre kostsamma jämfört med åtgärder att reducera smittspridningen långsiktigt). I Sverige var vi mycket snabba med att få ut vaccinet och vi har dessutom den högsta nationella vaccinationsnivån i världen, så andelen som smittades blev långt ifrån de ca 58% man kunde förvänta utan någon som helst prevention (se avsnitt 2.1). Nu när pandemin börjar lägga sig blir nästa stora uppgift att utvärdera vilka åtgärder som var mest effektiva och att jämföra olika länders åtgärder för att förbättra de preventiva metoderna till nästa gång. Ungefär hur många smittofall respektive dödsfall har Sverige haft respektive hur många har vi lyckats undvika genom att uppnå världens högsta vaccinationsgrad på ca 60%

(dock var den ju inte så hög när influensan anlände) och vad var kostnaden för detta?

## **Referenser**

Jag har medvetet inte givit en massa referenser till olika resultat. Vill man se aktuella tillämpningar kan man som sagt titta i aktuella nummer av t.ex. Nature och Science. Framträdande grupper finns bland annat vid Imperial College (Neil Ferguson), University of Washington (Ira Longini och Elizabeth Halloran) och Harvard (Mark Lipschits).

Vill man lära sig lite mer om modellerande av epidemier är nog boken *Infectious diseases of humans* av Anderson & May (Oxford, 1991) den överlägset mest inflytelserika. Jag kan även rekommendera *Mathematical epidemiology of infectious diseases* av Diekmann & Heesterbeek (Wiley, 2000) som håller en högre matematiskt nivå. Om man är intresserad av stokastiska modeller och statistiska inferensmetoder får jag vara ofin nog att referera till boken *Stochastic epidemic models and their statistical analysis* av Andersson & Britton (Springer, 2000). Mig veterligen finns bara en populärtext på svenska (utöver den ni har framför er!), även den skriven av undertecknad: "Pest, kolera och matematik", som ingår som ett kapitel i den alltför lite uppmärksammade boken *Människor och matematik – läsebok för nyfikna* (NCM, 2008). Jag rekommenderar inte boken för min egen artikel – speciellt inte när ni nu läst innevarande artikel – men det finns en mängd intressanta populärartiklar skrivna av ett 20-tal framstående svenska matematiker.