



FAKULTETEN FÖR
NATURVETENSKAPER OCH TEKNIK

AVHANDLING PRO GRADU

Epidemiologisk matematiska modeller

Skribent:

Minette BRINK

Handledare:

Mikael LINDSTRÖM

2021

Sammanfattning

I den här avhandlingen undersöks olika epidemiologiska matematiska modeller för att förutspå spridningen av infektionssjukdomar. Syftet med undersökningen är att analysera olika matematiska epidemiologiska modeller samt jämföra dem med modeller som använts för att uppskatta spridningen av COVID-19.

Det är många faktorer som påverkar och försvårar skapande av modellerna. Människors beteende går inte med exakt noggrannhet att förutspå, populationer går inte att begränsa noggrant och ovetskap om hur en infektionssjukdom sprids är faktorer som påverkar skapande av modellerna. I avhandlingen presenteras olika modeller var vi analyserar hur människors och infektionssjukdomars beteende påverkar ifall en epidemi bryter ut.

Reproduktionstalet är en viktig del inom matematiska epidemiologiska modeller. Det berättar väntade antalet personer en infekterad person smittar när alla andra personer i populationen är mottagliga för smitta. Beroende på modell och infektionssjukdomens beteende beräknas reproduktionstalet på olika sätt och kan därmed vara svårt att förutspå.

Kermacks och McKendricks [11] modell är en känd modell inom epidemiologisk matematisk modellering från 1927. Modellen som de skapade används fortfarande men med anpassningar för att vara exaktare. Modellen är *SIR* modellen,

$$\begin{cases} \frac{dS}{dt} = -\beta S \frac{I}{N} \\ \frac{dI}{dt} = \beta S \frac{I}{N} - \frac{1}{\tau} I \\ \frac{dR}{dt} = \frac{1}{\tau} I. \end{cases} \quad (1)$$

Modellen har relativt strikta antaganden och kan därmed sällan användas som sådan. Några anpassningar kan vara att personer inte blir immuna eller tillfrisknar inte av en sjukdom, sjukdomen har en latensperiod eller hur man har avgränsat populationen och personers beteende. Denna modell och anpassningar kommer att behandlas och analyseras i avhandlingen.

I och med att det är svårt att förutspå människors beteende och att avgränsa noggrant populationer är det svårt att estimeras spridningen av infektionssjukdomen. Modellen som används för att estimeras spridningen av COVID-19 har

anpassats separat för Italien, Spanien och USA [13] så att parametrar och data är lämpliga. I avhandlingen analyseras även resultaten av modellen. Fastän den är välformulerad ger det inte en garanti för att epidemin i verkligheten skulle bete sig som förutspått.

Innehåll

1	Inledning	4
1.1	Begrepp och beteckningar	5
1.2	Stabilitet för icke-linjära differentialekvationer	6
1.2.1	Linjär approximation och stabilitet	6
1.2.2	Stabilitet för icke-linjära system	7
2	Epidemi i en sluten population	9
3	Reproduktionstalet	11
3.1	Framställningen av reproduktionstal i enkla fall	12
3.2	Framställning av reproduktionstal i mer komplicerade fall	16
3.3	Nästa generationsmatris för indelad population	20
4	Första tillväxten av epidemin i realtid	27
5	Kermacks och McKendricks modell	30
5.1	Samarbete mellan Kermack och McKendrick	30
5.2	Beteckningar och definitioner	31
5.3	Matematiken i Kermacks och McKendricks modell	33
5.3.1	SIS modellen	36
5.3.2	SIER modellen	37
5.3.3	Exempel av SIR modellen	38
5.3.4	Smittsamhetstakt och väntetid	41
5.4	Kermack och McKendricks artikels framgång	44
6	Den slutliga storleken på en epidemi	45
6.1	Den slutliga storleken av en epidemi i en population utan restriktioner	46

6.2	Den slutliga storleken av en epidemi i en ändlig population	48
7	Modellering av COVID-19	50
7.1	Parametrarna under första vågen av COVID-19	53
7.1.1	Italien	54
7.1.2	Spanien	55
7.1.3	USA	55
7.2	Modellens beteende under den första, andra och eventuellt tredje vågen av COVID-19	56
7.2.1	Italien	56
7.2.2	USA	57
7.2.3	Spanien	58
7.3	Uppskattningar för utvecklingen av COVID-19 i Italien och USA .	58
7.3.1	Italien	59
7.3.2	USA	59
8	Avslutning	61

Tabeller

7.1	Tabell över värden för λ och μ år 2018 [13].	51
7.2	Kalibrering av γ för Italien [13].	55
7.3	Kalibrering av γ för USA [13].	56

Kapitel 1

Inledning

När ett virus introduceras, för en population, som kan bilda en endemi, epidemi eller pandemi är det vanligt att bilda matematiska modeller för att kunna förutspå virusets beteende. Det som påverkar modellerna är hur populationen är utspridd på ett område. Parametrar som bör beaktas i skapande av modeller är exempelvis i vilket område det finns åldringar, var det är glesbygd, om området är tätt bebott med mera. Som exempel finns det skillnader i hur ofta det förekommer fall av mässling när man jämför Storbritannien och Island. På Island förekommer det fler fall med en lägre frekvens än i Storbritannien och en av orsakerna är att Storbritannien är tätare befolkat än Island. En annan faktor som påverkar modellerna är hur många sjukdomsfall som inte rapporteras. Orsaken till att ett fall inte rapporteras kan vara att personen är symptomfri fastän hen har viruset, eller att hen inte låter sig testas. [4]

Institutet för hälsa och välfärd (THL) bildade i början på 2020 en grupp vars uppgift var att bilda matematiska modeller och scenarion för SARS-CoV-2 (COVID-19) och för att kunna bestämma restriktionsåtgärder som bör vidtas. Det som försvårade skapande av modellerna var att man inte visste hur länge personer är immuna eller om man får immunitet efter att ha tillfrisknat. En annan faktor som försvårade skapande av modellerna var att människors beteende kan vara svårt att förutspå. Det första som intresserade forskarna, efter att epidemin hade brutit ut i Wuhan, Kina, var reproduktionstalet som är förväntade antalet personer en infekterad person smittar när alla andra personer i populationen är mottagliga för smitta. Det går inte att räkna ut reproduktionstalet utifrån observationer utan det behövs matematiska modeller för att räkna ut

reproduktionstalet. [2]

I avhandlingen kommer centrala delar inom epidemiologisk matematiska modeller att behandlas. I kapitel 2 beskrivs hur stor betydelse populationen har och hur personers beteende kan påverka epidemiologiska modeller. Reproduktionstalet är en av de viktigaste delarna i epidemiologiska modeller. I kapitel 3 behandlas olika sätt att framställa reproduktionstalet beroende på infektionssjukdom och vilken modell man använt. Kapitel 4 handlar om första tillväxten av epidemin och en modell som beskriver början av en epidemi.

Kermack och McKendricks modell är en viktig modell inom matematisk modellering av infektionssjukdomar, som behandlas i kapitel 5. Där presenteras modellen som Kermack och McKendrick skapade, olika anpassningar av modellen samt några exempel. I kapitel 6 behandlas hur en epidemin ser ut i slutskedet beroende på om det har vidtagits åtgärder eller inte. Slutligen, i kapitel 7 ser vi en anpassning av Kermack och McKendricks modell där man har tillämpat data gällande COVID-19 från Spanien, Italien och USA.

1.1 Begrepp och beteckningar

I den här avhandlingen är terminologin främst från boken *Mathematical Tools for understanding Infectious Disease Dynamics* [6]. När andra källor används anpassas begreppen så att de är enhetliga genom hela avhandlingen.

I avhandlingen kommer det att antas att populationen är strukturerad, om inte annat nämns. En *strukturerad population* är en population som kan delas in i klasser eller grupper som ger en struktur till populationen.

En person som kan bli smittad, *mottaglig* person, är en person som kan bli smittad när hen är i kontakt med en infekterad person. Däremot är en *infekterad* person en som har blivit smittad och testar positivt för infektionen med hjälp av lämpligt prov. Svar som är fel i provresultaten kommer inte att beaktas för att göra modellerna enklare. En infekterad person behöver inte smitta andra under *inkubationstiden*. Inkubationstid är tidsintervallet mellan att en mottaglig person har blivit infekterad och sjukdomsutbrottet. Det antas att en person inte smittar under inkubationstiden för att göra modellerna enklare. En person som är *frisk* har varit infekterad och kan inte mera smitta personer som är mottagliga för smitta. Det antas även, i de flesta fall, att friska personer inte är mottagliga

för sjukdomen, åtminstone temporärt. Låt $i(t)$ beteckna *incidensen*, antalet nya infekterade över en tidsenhet.

Reproduktionstalet, betecknas R_0 , är förväntade antalet personer en infekterad person smittar när alla andra personer i populationen är mottagliga för smitta. Tröskelvärdet för R_0 är 1, när $R_0 > 1$ så ökar och när $R_0 < 1$ så minskar spridningen av infektionen i populationen. Alltså när $R_0 > 1$ så finns det en ökad risk att en epidemi bildas i en population.

En mottaglig person kommer att betecknas med S , en infekterad person med I och en frisk person med R . Dessa tre grupper bildar *SIR* modellen. Det finns även modeller som *SEIR*, där E är en person som blivit utsatt för smitta, och *SIS* där man inte kan bli immun mot infektionssjukdomen. Låt N beteckna hela populationen som beaktas och S, I, R är ofta angivna som andelar av N .

1.2 Stabilitet för icke-linjära differentialekvationer

I det här avsnittet behandlas stabilitet och approximation av icke-linjära system [10].

1.2.1 Linjär approximation och stabilitet

Betrakta

$$\frac{dx}{dt} = f(x) \quad (1.1)$$

som en autonom differentialekvation och $x(t)$ en obekant funktion. Funktionen $x = x_{ss}$ där $f(x) = 0$ kallas *jämviktspunkt* till (1.1). Taylors serieutveckling kring punkten x_{ss} ger

$$f(x) \approx f(x_{ss}) + \frac{df}{dx}(x_{ss})(x - x_{ss}) = \frac{df}{dx}(x_{ss})(x - x_{ss}) \quad (1.2)$$

där $f(x_{ss}) = 0$. Följaktligen kan (1.1) approximeras med

$$\frac{dx}{dt} \approx \frac{df}{dx}(x_{ss})(x - x_{ss}). \quad (1.3)$$

Låt $a = \frac{df}{dx}(x_{ss})$ och betrakta $\frac{dx}{dt} = a(x - x_{ss})$. Om $a < 0$, är den kritiska punkten x_{ss} *stabil*. Om $a > 0$ är den kritiska punkten x_{ss} *instabil*. Lösningen till

$\frac{dx}{dt} = ax - ax_{ss}$ ges av

$$x(t) = ce^{at} + x_{ss}, \quad c \in \mathbb{R}. \quad (1.4)$$

Följaktligen kommer avvikelsen från den kritiska punkten att minska exponentiellt om $a < 0$.

1.2.2 Stabilitet för icke-linjära system

Låt

$$\begin{cases} \frac{dx}{dt} = f(x, y) \\ \frac{dy}{dt} = g(x, y) \end{cases} \quad (1.5)$$

vara ett autonomt icke-linjärt system och $x(t)$, $y(t)$ obekanta funktioner. Punkterna (x, y) där både $f(x, y) = 0$ och $g(x, y) = 0$ kallas *jämviktsläge* till (1.5). Det är möjligt att man får ingen, ett ändligt antal eller oändligt många kritiska punkter. För att bestämma toppen och stabilitet av en kritisk punkt (x_{ss}, y_{ss}) approximerar vi (1.5) i en liten omgivning av (x_{ss}, y_{ss}) med ett linjärt system med konstanta koefficienter

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x - x_{ss} \\ y - y_{ss} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_{ss} \\ y_{ss} \end{pmatrix}, \quad (1.6)$$

där

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x}(x_{ss}, y_{ss}) & \frac{\partial f}{\partial y}(x_{ss}, y_{ss}) \\ \frac{\partial g}{\partial x}(x_{ss}, y_{ss}) & \frac{\partial g}{\partial y}(x_{ss}, y_{ss}) \end{pmatrix} \quad (1.7)$$

är jacobimatrisen. Denna matris bestämmer stabiliteten för den kritiska punkten (x_{ss}, y_{ss}) .

Bewis. Enligt Taylors serieutveckling kring punkten (x_{ss}, y_{ss}) fås

$$f(x, y) \approx f(x_{ss}, y_{ss}) + \frac{\partial f}{\partial x}(x_{ss}, y_{ss})(x - x_{ss}) + \frac{\partial f}{\partial y}(x_{ss}, y_{ss})(y - y_{ss}) \quad (1.8)$$

och

$$g(x, y) \approx g(x_{ss}, y_{ss}) + \frac{\partial g}{\partial x}(x_{ss}, y_{ss})(x - x_{ss}) + \frac{\partial g}{\partial y}(x_{ss}, y_{ss})(y - y_{ss}). \quad (1.9)$$

Emedan (x_{ss}, y_{ss}) är en kritisk punkt, gäller att $f(x_{ss}, y_{ss}) = 0$ och $g(x_{ss}, y_{ss}) = 0$. Således kan det icke-linjära systemet (1.5) approximeras med

$$\frac{dx}{dt} \approx \frac{\partial f}{\partial x}(x_{ss}, y_{ss})(x - x_{ss}) + \frac{\partial f}{\partial y}(x_{ss}, y_{ss})(y - y_{ss}) \quad (1.10)$$

och

$$\frac{dy}{dt} \approx \frac{\partial g}{\partial x}(x_{ss}, y_{ss})(x - x_{ss}) + \frac{\partial g}{\partial y}(x_{ss}, y_{ss})(y - y_{ss}) \quad (1.11)$$

eller

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x}(x_{ss}, y_{ss}) & \frac{\partial f}{\partial y}(x_{ss}, y_{ss}) \\ \frac{\partial g}{\partial x}(x_{ss}, y_{ss}) & \frac{\partial g}{\partial y}(x_{ss}, y_{ss}) \end{pmatrix} \begin{pmatrix} x - x_{ss} \\ y - y_{ss} \end{pmatrix}. \quad (1.12)$$

Låt $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ vara jacobimatrisen. Nu undersöks stabiliteten för den kritiska punkten. Detta beror på typen av egenvärden till A , det vill säga vilka är lösningarna λ_1, λ_2 av den karaktäristiska ekvationen

$$\det(A - \lambda I) = \det \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix} = \lambda^2 - (a + d)\lambda + \det A = 0. \quad (1.13)$$

Låt $p = a + d$, $q = \det A = ad - bc$ och $\Delta = p^2 - 4q$. Då följer att

$$\lambda^2 - p\lambda + q = (\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2. \quad (1.14)$$

Således är p summan av egenvärden, q är produkten av egenvärden och Δ är diskriminanten [12].

En kritisk punkt (x_{ss}, y_{ss}) kallas

- *asymptotiskt stabil* om $p < 0$ och $q > 0$,
- *stabil* om $p \leq 0$ och $q > 0$,
- *instabil* om $p > 0$ och $q < 0$.

Nu diskuteras kort dessa kriterier. Om $q = \lambda_1\lambda_2 > 0$, så är båda egenvärdena positiva, negativa eller komplexa konjugat. Om också $p = \lambda_1 + \lambda_2 < 0$, så är båda negativa eller har en negativ reell del. Då är (x_{ss}, y_{ss}) asymptotiskt stabil. Om $\Delta < 0$, är egenvärden komplexa konjugat, det vill säga $\lambda_1 = \alpha + i\beta$ och $\lambda_2 = \alpha - i\beta$. Om också $p = \lambda_1 + \lambda_2 = 2\alpha < 0$, så är (x_{ss}, y_{ss}) asymptotiskt stabil. Om $p = 0$, så är $\lambda_2 = -\lambda_1$ och $q = \lambda_1\lambda_2 = -\lambda_1^2$. Om också $q > 0$, så ger $\lambda_1^2 = -q < 0$, att λ_1 och därmed också λ_2 är rena imaginära tal. \square

Kapitel 2

Epidemi i en sluten population

Det här kapitlet är baserat främst på boken *Mathematical Tools for understanding Infectious Disease Dynamics* [6].

För att göra det enklare så antas det att populationen är sluten och att populationen inte utsätts för immigration eller emigration. Populationen ökar när barn föds och minskar när personer dör. Denna process sker under ett bestämt tidsintervall, till exempel tio år. När en smittsam sjukdom introduceras för en population varar sjukdomen kortare tid än exempelvis tio år i populationen, det kan vara frågan om så kort tid som några veckor. För att förenkla modelleringen beaktas inte, i de flesta fall, den naturliga ökningen och minskningen av den slutna populationen. Det antas även att det är en ny smittsam sjukdom som introduceras till populationen, populationen kallas då *naiv* för sjukdomen.

Det har betydelse hur populationen eller delpopulationen mäts. Det ska finnas en klar definition om hur populationen är mätt och man ska hålla sig till den för att analyserna ska vara konsekventa. För att jämföra två olika populationer är det viktigt att beakta hur mätningarna gjorts. Populationstätheten har också en betydelse. Beroende på vilken sjukdom som undersöks så kan tätheten ha en betydande roll, till exempel sjukdomar som sprids med droppsmitta sprids lättare då populationstätheten är hög.

Det är intressant att undersöka när en person från den naiva populationen blir smittad av en infektion om det kommer att skapa en epidemi och i så fall hur epidemin kommer att bete sig i populationen. Det antas även i vissa modeller att en person inte kan smittas två gånger av samma sjukdom, vilket underlättar uppskattande. Det betyder att personen blir immun efter att ha tillfrisknat eller

så dör hen.

För att en person ska bli smittad är det givet att personen måste komma i kontakt med infektionen. Olika infektioner smittar på olika sätt, malaria smittar via myggbett, HIV smittar bland annat via samlag och COVID-19 smittar via droppsmitta [17]. Det finns även infektioner var det är oklart hurdan kontakten ska vara för att smitta ska ske. För att förstå hur en infektion smittar skapas en modell över vilka personer som är smittade och vilka personer de har varit i kontakt med. Därefter skapas en modell över kontakterna mellan smittade personer och mottagliga personer. För att skapa modellerna antas det att en person har ett förväntat antal kontakter, c , per tidsenhet med andra personer.

Innan en epidemi bryter ut så finns det endast ett fåtal smittade personer i populationen och resten av populationen är mottaglig för smitta. Då kan det antas att en smittad person kommer i kontakt endast med mottagliga personer. Sannolikheten för att smitta ska ske, betecknas med p , beror på hur länge en smittad person har varit i kontakt med en mottaglig person. Det kan skrivas som

$$\begin{cases} 0, & \text{då } \tau < T_1 \\ p, & \text{då } T_1 \leq \tau \leq T_2 \\ 0, & \text{då } T_2 < \tau \end{cases} \quad (2.1)$$

där τ betecknar infektionens ålder, det vill säga hur lång tid det är sedan smitta inträffade. Låt $0 < p \leq 1$ och det antas att det finns en latenstid T_1 , den tiden då en person blivit smittad men ännu inte smittar andra. Under tidsintervallet $[T_1, T_2]$ smittar infekterade personer mottagliga personer. Efter T_2 blir smittade personer immuna eller så dör de. Om personen inte kan tillfriskna från sjukdomen så finns det exempelvis bromsmediciner som möjliggör att personen inte smittar. Om personer inte blir immun utan kan insjukna på nytt kan hen bli smittad på nytt med sannolikheten p .

Kapitel 3

Reproduktionstalet

Det här kapitlet kommer att handla om reproduktionstalet och är främst baserat på boken *Mathematical Tools for Understanding Infectious Disease Dynamics* [6] om inte annan källa nämns. Det finns flera olika sätt att räkna reproduktionstalet. Vi börjar med en enkel och begränsad modell och fortsätter sedan med mer komplicerade sätt.

Reproduktionstalet är en väsentlig del för att beskriva hur en epidemi beter sig. Det är den viktigaste delen för att skapa modeller för infektionssjukdomar och är bland det första som estimeras när en epidemi bryter ut. Dock är det inte entydigt att estimeras. När man har estimerat reproduktionstalet kan det tas nödvändiga åtgärder som hindrar att en stor epidemi bryter ut. I Finland har myndigheterna gett direktiv för olika åtgärder och restriktioner för att minska på fallen av COVID-19 som till exempel att stänga restauranger, skolor och bibliotek. [18].

Definition 3.1. *Reproduktionstalet är förväntade antalet personer en infekterad person smittar när alla andra personer i populationen är mottagliga för smitta. Reproductionstalet betecknas med R_0 .*

Reproduktionstalet är ett relativt nytt koncept inom epidemiologi. Det härstammar från demografi och ekologi var reproduktionstalet beskrev genomsnittliga antal flickor födda av kvinnor. Den första som introducerade R_0 var Richard Böckh år 1886. Han använde en tabell år 1879 som beskrev sannolikheten för att kvinnor överlever vid födsel av flickor. Han kom fram till att i genomsnitt kommer 2,172 flickor att födas av varje kvinna. Varpå han räknade R_0 talet att vara 1,06. [14]

3.1 Framställningen av reproduktionstal i enkla fall

I enkla fall kan reproduktionstalet räknas på följande sätt,

$$R_0 = pc(T_2 - T_1) \quad (3.1)$$

där p , T_1 och T_2 är samma som i (2.1) och c är förväntade antalet kontakter mellan en infekterad och en mottaglig person per tidsenhet.

Tröskelvärdet för R_0 är 1, när $R_0 > 1$ så ökar spridningen av infektion och när $R_0 < 1$ så minskar spridningen av infektionen i populationen. Alltså när $R_0 > 1$ så finns det en ökad risk att en epidemi bryter ut i en population.

För att en epidemi ska bryta ut när $R_0 > 1$ så behövs det ett stort antal både smittade och mottagliga personer i populationen. För att framställa R_0 rätt är demografisk slumpmässighet ett viktigt antagande. Andelen av populationen som är infekterade beaktas och inte antalet infekterade personer. Därför måste en liten andel av populationen vara smittad för att beräkna R_0 .

För att förenkla infektionsspridning jämförs det med förökning i en population. Populationens ökning från generation till generation är i modeller en liknande händelse som infektionsspridning. Det antas att en liten andel av populationen är infekterad för att infektionen ska börja spridas och att i början av epidemin är andelen mottagliga personer i populationen stor så att vi kan förbise de personer som blir infekterade i början.

Det antas att personer i en population förökar sig oberoende av varandra och antalet avkomor är fördelade enligt fördelningen $\{q_k\}_{k=0}^{\infty}$. En person föder k stycken barn med sannolikheten q_k och $\sum_{k=0}^{\infty} q_k = 1$. Från det kan utläsas att väntade antal barn är R_0 , som kan skrivas som

$$R_0 = \sum_{k=0}^{\infty} kq_k. \quad (3.2)$$

För att använda det här mera generellt skrivs det om till sannolikhetsgenererande funktionen

$$g(z) = \sum_{k=0}^{\infty} q_k z^k, \quad 0 \leq z \leq 1. \quad (3.3)$$

Det antas att $q_0 > 0$, det betyder att det finns en positiv sannolikhet att en person inte får någon avkomma. För att börja enkelt beaktas först en enpersons

population. Då dör populationen ut med sannolikheten q_0 efter ett generationssteg. Låt z_n beteckna sannolikheten för att en population dör ut efter n generationer. Då är $z_0 = q_0$ och vidare $z_n = g(z_{n-1})$. I den första generationen föds k stycken personer. För att populationen ska dö ut innan generation n ska varje generation ända till $n - 1$ ha dött som härstammar från de k stycken personer i första generationen. Sannolikheten för att det ska ske vid varje gren av generationer från k personerna är z_{n-1} . Sannolikheterna är oberoende så vi får att sannolikheten för att samtliga grenar av generationer från de k personerna dör ut är z_{n-1}^k . Det kvarstår att summera för alla möjliga värden av k med lämpliga vikten q_k , alltså

$$z_n = q_0 + \sum_{k=1}^{\infty} q_k z_{n-1}^k = g(z_{n-1}). \quad (3.4)$$

Låt g vara växande vilket betyder att z_n är växande (se beviset av sats 3.2), z_n är begränsad av 1 så gränsvärdet existerar och är $z_\infty = \lim_{n \rightarrow \infty} z_n$. Enligt hur z_n definierats är z_∞ sannolikheten för att populationen påbörjad av den första personen kommer att dö ut. Denna beskrivning går att jämföra med ett mindre utbrott. När $z_\infty = 1$ kommer populationen att dö ut med sannolikheten 1. När $0 < z_\infty < 1$ existerar komplementet som är $1 - z_\infty$. Alltså $R_0 < 1$ när $z_\infty = 1$ och $R_0 > 1$ då $0 < z_\infty < 1$. Även om $R_0 > 1$ så kan sjukdomen dö ut av olika orsaker då antalet infekterade är lågt.

Sats 3.2. *Sannolikheten att epidemin dör ut i ett tidigt skede fastän $R_0 > 1$ fås från att z_∞ är den minsta roten i $[0, 1]$ till ekvationen $z = g(z)$.*

Bevis. Anta att $q_0 > 0$ och därmed är $g(0) = q_0 > 0$. På grund av att g är växande, så är $g(q_0) > g(0) = q_0$ eller $z_1 > z_0$. Det här implicerar att $z_2 = g(z_1) > g(z_0) = z_1$. Genom induktion fås att z_n växer. För att $g(z) \leq g(1) = 1$ är följden begränsad. En begränsad och växande följd har ett gränsvärde, som i det här fallet kallas för z_∞ . Genom att beräkna gränsvärdet då $n \rightarrow \infty$ då $z_n = g(z_{n-1})$ fås $z_\infty = g(z_\infty)$, då g är kontinuerlig. Låt w beteckna den minsta lösningen till ekvationen $z = g(z)$. Då $w > 0$ och därmed är $w = g(w) > 0 = q_0 = z_0$. Genom att applicera g upprepade gånger fås att $w > z_n$ för alla n . Då $n \rightarrow \infty$ fås att $w \geq z_\infty$. Men för att w definierades som den minsta roten så måste det vara likhet, alltså $w = z_\infty$. \square

Tolkningen av villkoret för hållbarhet är följande. Sannolikheten att dö ut är lika med sannolikheten att dö ut genast adderat med, summan över k , sanno-

likheten att dö ut betingat att infektera k andra personer. Matematisk uttryckt är

$$z_\infty = q_0 + \sum_{k=1}^{\infty} q_k z_\infty^k = g(z_\infty) \quad (3.5)$$

där det har använts oberoendehet för att skriva sannolikheten att alla k grenar dör ut vid z_∞^k .

Under tidsintervaller ΔT följer mängden kontakter poissonfördelningen $X \sim \text{Po}(k)$, $k > 0$ med parametern c . Fördelningen beskriver sannolikheten att kontakterna k är jämnt fördelade

$$P(x = k) = \frac{(c\Delta T)^k}{k!} e^{-c\Delta T}, \quad k = 0, 1, 2, \dots \quad (3.6)$$

Antalet kontakter följer poissonfördelningen med medelvärdet $c\Delta T$.

Sats 3.3. När kontakterna leder till smitta med sannolikheten p är antalet fall av smitta även poissonfördelat med parametern $pc\Delta T$.

Bevis. Om k personer har haft kontakt och m av dem är infekterade personer är sannolikheten

$$\binom{k}{m} p^m (1-p)^{k-m}, \quad (3.7)$$

där $\binom{k}{m} = \frac{k!}{m!(k-m)!}$. Därmed är sannolikheten för att m personer är infekterade

$$\begin{aligned} & \sum_{k=m}^{\infty} \binom{k}{m} p^m (1-p)^{k-m} \frac{(c\Delta T)^k}{k!} e^{-c\Delta T} \\ &= \frac{p^m}{m!} (c\Delta T)^m e^{-c\Delta T} \sum_{k=m}^{\infty} \frac{(1-p)^{k-m}}{(k-m)!} (c\Delta T)^{k-m} \\ &= \frac{p^m}{m!} (c\Delta T)^m e^{-c\Delta T} \sum_{l=0}^{\infty} \frac{(1-p)^l}{l!} (c\Delta T)^l \\ &= \frac{p^m}{m!} (c\Delta T)^m e^{-c\Delta T} e^{(1-p)c\Delta T} \\ &= \frac{(pc\Delta T)^m}{m!} e^{-pc\Delta T} \end{aligned} \quad (3.8)$$

□

Genererande fördelningsfunktionen för poissonfördelningen med parametern λ är

$$\begin{aligned} g(z) &= \sum_{k=0}^{\infty} z^k \frac{\lambda^k e^{-\lambda}}{k!} = \sum_{k=0}^{\infty} \frac{(\lambda z)^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} e^{\lambda z} \sum_{k=0}^{\infty} \frac{(\lambda z)^k e^{-\lambda z}}{k!} = e^{\lambda(z-1)}. \end{aligned} \quad (3.9)$$

Där $\frac{\lambda^k e^{-\lambda}}{k!} = q_k$ och $\sum_{k=0}^{\infty} \frac{(\lambda z)^k e^{-\lambda z}}{k!} = 1$, emedan termerna i summan är sannolikheten då $\text{Po}(\lambda z)$ blir 1. Den första likheten fås från definitionen för genererande fördelningsfunktionen för poissonfördelningen. Tillsammans med att $R_0 = pc\Delta T = pc(T_2 - T_1) = \lambda$ fås att z_∞ då $R_0 > 1$ är en entydig rot i intervallet $(0, 1)$ för ekvationen

$$z = e^{R_0(z-1)}. \quad (3.10)$$

För att undvika missförståndet att det här gäller generellt lägger vi till sannolikheten för mindre utbrott $z_\infty = \frac{1}{R_0}$ för en annan delmodell för smittsamhet, där ΔT inte är fixerat utan en godtycklig variabel som följer från exponentiella fördelningen.

Sats 3.4. *Låt ΔT vara exponentiellt fördelad med parametern α . Då gäller*

$$\begin{aligned} g(z) &= \alpha \int_0^\infty e^{pc\Delta T(z-1)} e^{-\alpha\Delta T} d(\Delta T) \\ &= \frac{\alpha}{\alpha - pc(z-1)}. \end{aligned} \quad (3.11)$$

Bevis. För ett givet ΔT är sannolikheten för att m personer blir infekterade $\frac{(pc\Delta T)^m}{m!} e^{-pc\Delta T}$. Enligt sats 3.3 följer att sannolikhetstäthetsfunktionen för ΔT är $\alpha e^{-\alpha\Delta T}$ med $\alpha > 0$. Därmed fås

$$q_m = \alpha \int_0^\infty e^{-\alpha\Delta T} \frac{(pc\Delta T)^m}{m!} e^{-pc\Delta T} d(\Delta T) \quad (3.12)$$

och

$$g(z) = \alpha \int_0^\infty e^{-\alpha\Delta T} e^{pc\Delta T(z-1)} d(\Delta T). \quad (3.13)$$

Emedan $g(z)$ är en potensserie är omskrivningen mellan summan och integralen i nedanstående beräkningen giltig,

$$\begin{aligned} g(z) &= \sum_{m=0}^{\infty} z^m \alpha \int_0^\infty e^{-\alpha\Delta T} \frac{(pc\Delta T)^m}{m!} e^{-pc\Delta T} d(\Delta T) \\ &= \alpha \int_0^\infty \left[\sum_0^\infty z^m \frac{(pc\Delta T)^m e^{-pc\Delta T}}{m!} \right] e^{-\alpha\Delta T} d(\Delta T) \\ &= \alpha \int_0^\infty e^{pc\Delta T(z-1)} e^{-\alpha\Delta T} d(\Delta T), \end{aligned} \quad (3.14)$$

varför (3.13) gäller.

När man räknar ut integralen fås

$$g(z) = \frac{\alpha}{\alpha - pc(z-1)}. \quad (3.15)$$

Nu undersöks det ännu att $z_\infty = \frac{1}{R_0}$ stämmer. Vi har ekvationen $z = \frac{\alpha}{\alpha - pc(z-1)}$ som har rötterna $z = 1$ och $z = \frac{\alpha}{pc}$. Det finns endast två rötter alltså ekvationen stämmer. \square

3.2 Framställning av reproduktionstal i mer komplicerade fall

I det här avsnittet kommer det att visas ett annat sätt för att framställa reproduktionstalet. Det antas att en infektion har ändligt antal tillstånd $1, 2, \dots, n$. För att bestämma R_0 fås följande induktionsproblem: anta en population med demografisk stadigt tillstånd. Populationen kommer i kontakt med en ny infektion som populationen inte har någon historia av. Kommer då den nya infektionen att skapa ett utbrott i populationen?

För att förenkla används det deterministiska synsättet, med specifikt indata får vi samma utdata, och systemet görs linjärt. Alltså beaktas inte att infektionen i sig själv minskar antalet personer som är mottagliga för smitta. Antagande om demografiskt stadigt tillstånd betyder att infektionen har en miljö som är konstant.

Reproduktionstalet är matematiskt karakteriserat som en demografisk process. För att kunna jämföra generationer i en demografisk process med generationer av smitta kommer avkomman att vara en ny infektion orsakad av smitta, så kallad epidemiologisk födsel. När en generation ökar i storlek betyder det att epidemin växer, alltså indikerar tillväxtfaktorn i generationer på en potentiell tillväxt i epidemin. Det betyder att R_0 motsvarar tillväxtfaktorn.

Väntevärdet för nya fall som har infektionstillstånd i orsakad av en person med infektionstillstånd j , under tiden då personen i tillstånd j är smittsam, där alla kontakter med personen i tillstånd j är med mottagliga personer betecknas med k_{ij} . Med denna definition av k_{ij} fås n^2 icke-negativa tal. Det intressanta är att söka ett värde med två följande egenskaper, introduktionen av infektionen lyckas om och endast om värdet är större än 1 och värdet har en biologisk tolkning liknande till ett homogent fall. För att klargöra det här presenteras följande exempel.

Exempel 3.5. Låt $K = (k_{ij}) = \begin{pmatrix} 0 & 100 \\ 10 & 0 \end{pmatrix}$ som kan beskriva en infektion som

sprids mellan djur och människor, som har ett djur och människotillstånd av infektion. Ett djur i tillstånd 1 skapar i medeltal 10 nya fall med infektionstillstånd 2. Varje person i infektionstillstånd 2 skapar i medeltal 100 nya fall med infektionstillstånd 1. Eftersom ett djur inte kan smitta ett annat djur, utan djuret smittar enbart människor som i sin tur enbart smittar djur, alternerar faktorerna 10 och 100. Alltså tar det två generationer för infektionen att åter bli i ursprungstillstånd. Varje generationstal multipliceras med $10 \cdot 100 = 1000$. Då fås varje generations multiplikationsfaktor $\sqrt{1000}$.

Generationer av smittade personer kan beskrivas med vektorer. Det j te elementet av vektorn är antalet fall vid infektionstillstånd j i den specifika generationen. Vektorn för nästa generation fås från vektorn som beskriver nuvarande generationen multiplicerat med matrisen K , som kallas för nästa generationsmatris. Alltså

$$\phi^{ny} = \sum_{j=1}^n k_{ij} \phi_j^{gammal} \quad (3.16)$$

där ϕ^{ny} är antalet personer i mottagligt tillstånd i som är smittade av en person med infektionstillstånd j , summerade med totala tillstånd j . Då fås

$$\phi^{ny} = K \phi^{gammal} \Rightarrow \phi^{m+1} = K \phi^m. \quad (3.17)$$

Låt $K^m = K \cdot \dots \cdot K$ (m stycken K), så gäller likheterna

$$\phi^m = K \phi^{m-1} = K^2 \phi^{(m-2)} = \dots = K^m \phi^0 \quad (3.18)$$

där ϕ^0 är generation 0 som börjar processen. Vidare är K är en icke-negativ matris, alltså $K = (k_{ij}) > 0$, $i > 1$ och $n \geq j$ och även vektorerna $\phi \geq 0$ är icke-negativa.

Den naturliga normen för vektorer, som mäter vektorers storlek, är

$$\|\phi\| = \sum_{j=1}^n |\phi_j|, \quad (3.19)$$

där $\phi = (\phi_1, \dots, \phi_n)^\top$. Alltså är $\|\phi\|$ totala antalet fall i en generation som beskrivs med ϕ . Normen av matrisen K beskrivs som vanligt med

$$\|K\| = \sup_{\|\phi\| \neq 0} \frac{\|K\phi\|}{\|\phi\|} = \sup_{\|\phi\|=1} \|K\phi\| \quad (3.20)$$

Inom epidemiologi kan $\|K\|$ tolkas som den maximala multipeln för det totala antalet fall, när fördelningen med avseende på stadiet av infektionen kan ta alla möjliga former. Från det tidigare exemplet 3.5 fås att $\|K\| = 100$. Det här betyder att ϕ^0 kan väljas godtyckligt men att fördelningen bestäms av dynamiken. I exempel 3.5 är $\phi^0 = \begin{pmatrix} 0 & 10 \end{pmatrix}^\top$ och $\phi^1 = \begin{pmatrix} 100 & 0 \end{pmatrix}^\top$ är antalet fall multiplicerade med 100. Däremot multipliceras antalet fall med 10 i nästa generation. Följaktligen är normen av K för grov att bestämma nästa generations tillväxt. Därför undersöks multiplikationen i m generationer för varje generation. Alltså är exponenten $1/m$ av tillväxten i m generationer intressant i $\|K^m\|^{1/m}$.

Generellt behöver man ha tålamod och beakta $m \rightarrow \infty$. Det här gränsvärdet kallas för spektralradien, och betecknat med $\rho(K)$. Enligt definitionen är spektralradien det långsiktiga medelvärdet per generationens multiplikationstal. Från det här fås karaktäriseringen av R_0

$$R_0 = \rho(K) = \lim_{m \rightarrow \infty} \|K^m\|^{1/m} \quad (3.21)$$

där K är nästa generationsmatris. Då är spektralradien R_0 ett egenvärde av K som kallas dominerande egenvärdet där $|\lambda| \leq R_0$ för alla egenvärden λ av K .

Det antas att R_0 är strikt dominant, alltså $|\lambda| < R_0$ för alla egenvärden λ (som är olika R_0) av K och att R_0 är ett algebraiskt enkelt egenvärde. Det betyder att det finns endast en faktor $\lambda - R_0$ i den karaktäristiska ekvationen $\det(\lambda I - K) = 0$. Egenvektorn ψ^d (d står för dominerande) för R_0 kan väljas så att alla dess komponenter är icke-negativa och att deras summa är lika med ett, alltså är ψ^d normaliserad. Från det här kan det visas att

$$K^m \phi^0 = c(\phi^0) R_0^m \psi^d + o(R_0^m) \text{ för } m \rightarrow \infty \quad (3.22)$$

där $o(R_0^m)$ uttrycker att $\frac{[K^m \phi^0 - c(\phi^0) R_0^m \psi^d]}{R_0^m} \rightarrow 0$ då $m \rightarrow \infty$.

Den normaliserade dominanta egenvektorn, ψ^d , beskriver den stabiliserade distributionen. Då fås två egenskaper, invarians och attraktion. Om $\phi^0 = \psi^d$ fås att $K^m \psi^d = R_0^m \psi^d$, då är fördelningen oförändrad från generation till generation när talen multipliceras med R_0 . För generella förstagersvektorer liknar situationen mer och mer det här specialfallet. Endast skalären $c(\phi^0)$ hålls oförändrad.

När R_0 är strikt dominerande, algebraiskt enkelt, är ϕ^0 sådant att $c(\phi^0) \neq 0$. Då är det klart att när nästa generationsmatris, K , appliceras upprepade gånger

kommer den normaliserade fördelningen att konvergera mot den stabila fördelningen ψ^d . Därtill kommer generationsmultipeln att konvergera mot R_0 . Den normaliserade stabila fördelningen kan beskrivas som sannolikhetsfördelningen för stadiet av infektion.

Definition 3.6. Den icke-negativa matrisen K är oreducerbar om för varje indexpar i och j finns ett heltal $m = m(i, j) > 0$ så att $(K_{ij}^m) > 0$. K är primitiv om det finns ett m för alla i och j , alltså om det finns ett m så att alla element i K^m är strikt positiva.

En oreducerbar matris är en matris vars rader och kolumner inte kan permuteras så att den omvandlas till övre blockets triangelform.

Följande slutsatser kan dras om K är primitivt.

- R_0 är strikt dominerande
- ψ^d och ψ^{d*} har strikt positiva komponenter (och som en konsekvens $c(\phi^0) > 0$ för vilken som helst icke trivialt positivt ϕ^0)
- R_0 är ett algebraiskt enkelt egenvärde
- Inga andra egenvärden har positiva egenvektorer.

När K är nästan oreducerbar, håller de tre sista egenskaperna, men på cirkeln av radien R_0 i komplexa talplanet finns det andra egenvärden. Dessa egenvärden är rötter från ekvationen $\lambda^m = R_0^m$ för något m , alltså måste de vara i formen $\lambda = R_0 \exp(i\frac{2\pi l}{m})$ där $l = 1, \dots, m - 1$.

Det viktiga antagande är att antalet personer som är involverade i förändringen är stort. Då fås skarpa tröskelvärden, $R_0 < 1$ som inte leder till epidemi och om $R_0 > 1$ leder det till en epidemi som påverkar andelen av populationen som blivit infekterad. I verkligheten är tröskelvärdet inte lika skarpt som i modellerna. Det beror på att det är osäkert om en mottaglig person blir smittad när hen kommer i kontakt med en smittad person. Även om $R_0 > 1$ så kan det leda till ett mindre utbrott och fastän R_0 är lite mindre än 1 så kan det leda till ett större utbrott. Det är vanligt att i stället för att bestämma slutliga storleken av epidemin så bestäms en sannolikhetsfördelning för den slutliga storleken av epidemin.

3.3 Nästa generationsmatris för indelad population

Tidigare definierades R_0 med hjälp av generationer som leder till karaktärisering av önskade egenskaper från homogena till heterogena system. I den här delen kommer det att behandlas hur generationsbeskrivningen av infekterade personer är relaterad till en realtidsbeskrivning. Det kommer även att presenteras tre olika nästa generationsmatriser, en av dem behandlades redan tidigare, K .

I indelade modeller kan personers infektioner vara diskretuppdelade i en ändlig mängd stadier. Mängden av personer kan delas i delmängder, som kan beskriva olika stadier av sjukdomen. Tidigare definierades R_0 , som spektralradien av nästa generationsmatris K , som är baserad på alla möjliga delmängder i systemet. Alltså R_0 är stadiet av sjukdomen som en person har genast efter att hen blivit infekterad. Det är vanligt att en sjukdom har flera stadier än stadiet då man insjuknar. Dessa stadier beaktas exempelvis i *SEIR* modellen. I *SEIR* modellen finns det två stadier då en person blir infekterad, E betecknar latenstiden och I betecknar infektionsstadiet. I det här fallet är det endast E som är stadiet av infektion för alla personer som blir smittade. Personen som blivit smittad börjar med stadiet E , alltså en person kan inte genast komma till stadiet I efter smitta. I det här fallet är nästa generationsmatrisen, K , en endimensionell matris. Man kan även bilda en tvådimensionell matris genom att bilda skilda generationsmodeller både för E och I .

För att beräkna R_0 behövs ekvationer som finns i systemet med ordinära differentialekvationer

$$\begin{cases} \frac{dS}{dt} = -\beta S \frac{I}{N} \\ \frac{dI}{dt} = \beta S \frac{I}{N} - \frac{1}{\tau} I \\ \frac{dR}{dt} = \frac{1}{\tau} I. \end{cases} \quad (3.23)$$

Systemet beskriver uppkomst av ny infektion och när infektionen ändrar stadie bland infekterade personer. Ekvationssystemet kommer att kallas infektionsdel-system. Först ska infektionsdelsystem göras linjärt som består av icke-linjära ekvationer som handlar om infektionsfria stadiga stadiet, som i regel finns (det behandlas närmare i avsnitt 1.2 och kapitel 5). Att göra systemet linjärt betyder epidemiologiskt att R_0 beskriver den potentiella spridningen av en population när den introduceras för en mottaglig population. Det antas även att popula-

tionen är tillräckligt stor så att andelen mottagliga i populationen inte ändras märkbart i början av spridningen.

Låt x vara vektorn som beskriver storleken på den infekterade delpopulationen. Låt det linjära infektionsdelsystemet vara i formen $\frac{dx}{dt} = (T + \Sigma)x$. Matrisen T beskriver smitta och matrisen Σ beskriver förändring i stadiet av infektion (till exempel när en person går från mottaglig till infekterad) och när personer tillfrisknar eller dör efter att varit infekterade. Alla epidemiologiska händelser som leder till en ny infektion hör till T och alla händelser i infektionsstadierna hör till Σ . Elementen i matrisen T betecknas med T_{ij} som är graden då en infekterad person i stadie j , lägger till en annan person till infektionsstadiet i . Elementen i icke-negativa matrisen $-\Sigma^{-1}$ som är den väntade tiden en person kommer att tillbringa i stadiet i givet att hen är i stadie j . Från det här kan det definieras en annan nästa generationsmatrix

$$K_S = -T\Sigma^{-1}, \quad (3.24)$$

som kommer att kallas *nästa generationsmatrix med stor värdemängd*. Elementen i $(K_S)_{ij}$ beskriver det väntade antalet nya infektioner börjande från stadie i , orsakad av en infekterad person i stadie j . Generellt gäller inte att K_S är lika med nästa generationsmatrix K . Vanligtvis har K_S större dimension än K , det här beror på att K_S beskriver generationer av infekterade personer uppdelade i de olika infektionsstadierna. Medan nästa generationsmatrix K är definierad så att den beskriver endast indelningen av infektionen.

Man kan beräkna K_S med matrisbeskrivningen av det linjära infektionsdelsystemet. Denna matrix bestämmer utvecklingen av generationerna av personer i alla infektionsstadier. Nästa generationsmatrix K bestäms av delmängden av dessa stadier och motsvarar K_S . Från epidemiologisk synvinkel kan infektionsstadier som också är stadier av infektion skiljas från de stadier som inte är det. Det här går att observera från T .

Exempel 3.7. *I det här exemplet beaktas följande stadier:*

- S : mottaglig
- E_1 latenstid i kategori 1
- E_2 latenstid i kategori 2

- I : Infekterad och smittsam
- R : tillfrisknad och därmed immun eller död på grund av sjukdomen
- p och $(1 - p)$: två olika kategorier i population
- β : graden för smitta
- μ : graden av födsel och dödlighet per capita
- v_1 och v_2 : graden för att lämna respektive latensgrupp och gå in i smittsamma stadiet
- α : graden för att lämna smittsamma stadiet

Den linjäriserade modellen för infekterade delsystemet ges av

$$\begin{cases} \frac{dE_1}{dt} = p\beta I - (v_1 + \mu)E_1 \\ \frac{dE_2}{dt} = (1 - p)\beta I - (v_2 + \mu)E_2 \\ \frac{dI}{dt} = v_1 E_1 + v_2 E_2 - (\alpha + \mu)I \end{cases} \quad (3.25)$$

med matriserna

$$T = \begin{pmatrix} 0 & 0 & p\beta \\ 0 & 0 & (1 - p)\beta \\ 0 & 0 & 0 \end{pmatrix} \quad (3.26)$$

och

$$\Sigma = \begin{pmatrix} -(v_1 + \mu) & 0 & 0 \\ 0 & -(v_2 + \mu) & 0 \\ v_1 & v_2 & -(\alpha + \mu) \end{pmatrix}. \quad (3.27)$$

Nästa generationsmatris med stor värdemängd K_S blir

$$\begin{aligned} K_S &= -T\Sigma^{-1} = \begin{pmatrix} 0 & 0 & p\beta \\ 0 & 0 & (1 - p)\beta \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{v_1 + \mu} & 0 & 0 \\ 0 & \frac{1}{v_2 + \mu} & 0 \\ \frac{v_1}{(v_1 + \mu)(\alpha + \mu)} & \frac{v_2}{(v_2 + \mu)(\alpha + \mu)} & \frac{1}{(\alpha + \mu)} \end{pmatrix} \\ &= \begin{pmatrix} \frac{p\beta v_1}{(v_1 + \mu)(\alpha + \mu)} & \frac{p\beta v_2}{(v_2 + \mu)(\alpha + \mu)} & \frac{p\beta}{\alpha + \mu} \\ \frac{(1 - p)\beta v_1}{(v_1 + \mu)(\alpha + \mu)} & \frac{(1 - p)\beta v_2}{(v_2 + \mu)(\alpha + \mu)} & \frac{(1 - p)\beta}{\alpha + \mu} \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \quad (3.28)$$

Låt R_0 vara det dominanta egenvärdet för K_S som fås av den karaktäristiska ekvationen

$$\lambda^2 - \lambda \left(\frac{p\beta v_1}{(v_1 + \mu)(\alpha + \mu)} + \frac{(1-p)\beta v_2}{(v_2 + \mu)(\alpha + \mu)} \right) = 0. \quad (3.29)$$

Då fås

$$R_0 = \left(\frac{pv_1}{v_1 + \mu} + \frac{(1-p)v_2}{v_2 + \mu} \right) \frac{\beta}{\alpha + \mu}. \quad (3.30)$$

Matrisen ϵ består av enhetsvektorer e_i för alla i så att raden i i T är olika noll. Då är ϵ en matris vars kolumner består av enhetsvektorer som relaterar till rader som är olika noll i T . I det här exemplet fås

$$\epsilon = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}. \quad (3.31)$$

Då består K av

$$\begin{aligned} K &= \epsilon^\top K_S \epsilon = -\epsilon^\top T \Sigma^{-1} \epsilon. \quad (3.32) \\ K &= \begin{pmatrix} 0 & 0 & p\beta \\ 0 & 0 & (1-p)\beta \end{pmatrix} \begin{pmatrix} \frac{1}{v_1 + \mu} & 0 \\ 0 & \frac{1}{v_2 + \mu} \\ \frac{v_1}{(v_1 + \mu)(\alpha + \mu)} & \frac{v_2}{(v_2 + \mu)(\alpha + \mu)} \end{pmatrix} \\ &= \begin{pmatrix} \frac{p\beta v_1}{(v_1 + \mu)(\alpha + \mu)} & \frac{p\beta v_2}{(v_2 + \mu)(\alpha + \mu)} \\ \frac{(1-p)\beta v_1}{(v_1 + \mu)(\alpha + \mu)} & \frac{(1-p)\beta v_2}{(v_2 + \mu)(\alpha + \mu)} \end{pmatrix} \end{aligned} \quad (3.33)$$

Det dominerande egenvärdet för K fås av den karaktäristiska ekvationen

$$\lambda^2 - \lambda \left(\frac{p\beta v_1}{(v_1 + \mu)(\alpha + \mu)} + \frac{(1-p)\beta v_2}{(v_2 + \mu)(\alpha + \mu)} \right) = 0, \quad (3.34)$$

som är samma som för K_S (3.30).

Den sista raden i matrisen T , i exemplet 3.7, består endast av nollor, det betyder att personer inte genast kan komma i stadie I efter att de blivit smittade. Alltså har systemet endast två stadier av infektion. Alla personer börjar sjukdomen i antingen stadie E_1 eller E_2 . Därför är nästa generationsmatris K tvådimensionell och K_S är tredimensionell.

Betrakta $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, där $a, b, c, d \in \mathbb{R}^+$. Spåret av A , $\text{tr}(A)$, definieras som $\text{tr}(A) = a + d$. Låt λ_1 och λ_2 vara egenvärden till A . Då ges spektralradien av A genom

$$\rho(A) = \max\{|\lambda_1|, |\lambda_2|\}. \quad (3.35)$$

Det gäller att $\det(\lambda I - A) = \det \begin{pmatrix} \lambda - a & b \\ c & d - \lambda \end{pmatrix} = 0$, det vill säga $(\lambda - a)(\lambda - d)bc = \lambda^2 - (a + d)\lambda + ad - bc = 0$. Här av fås

$$\begin{aligned} \rho(A) &= \frac{a + d + \sqrt{(a + d)^2 - 4(ad - bc)}}{2} \\ &= \frac{\text{tr}(A) + \sqrt{\text{tr}(A)^2 - 4 \det(A)}}{2}. \end{aligned} \quad (3.36)$$

Alltså det dominerande egenvärdet till K fås som

$$R_0 = \rho(K) = \frac{1}{2} \left[\text{tr}(K) + \sqrt{\text{tr}(K)^2 - 4 \det(K)} \right]. \quad (3.37)$$

I exempel 3.7 är $\det(K) = 0$, alltså K är en singular matris. På grund av att K är en 2×2 matris fås $R_0 = \text{tr}(K)$. Förutom att det är enkelt att hitta R_0 i en 2×2 matris finns det andra fördelar med att $\det(K) = 0$. Då $\det(K) = 0$ kan vi minska på dimensionen i matrisen. Det här är vanligt när incidensen motsvara två eller då flera olika stadier av infektion förekommer oberoende av tid. Med det menas att det finns stokastiskt endast ett stadie av infektion även om det egentligen finns två olika stadier av infektion. Alltså är sannolikhetsfördelningen över stadie av infektion fixerad, den epidemiologiska tolkningen är alltså att smitta inte är infekterade personers ansvar.

När dimensionerna minskas på K fås en matris som kallas *nästa generationsmatris med liten värdemängd* och betecknas med K_L . Om minskning av dimensionerna i K är möjligt, bestäms K_L från K . För att komma fram till K_L undersöks raderna och kolumnerna i T . I exempel 3.7 har T två kolumner bestående endast av nollor och en kolumn som är olika noll. För att framställa kolumnerna i T multipliceras varje kolumn med vektorn $C = \begin{pmatrix} p & 1 - p & 0 \end{pmatrix}^\top$. De två första kolumnerna är C multiplicerat med noll och den sista kolumnen är C multiplicerat med β . För att framställa raderna i T multipliceras raderna på motsvarande sätt med vektorn $R = \begin{pmatrix} 0 & 0 & \beta \end{pmatrix}$. Den första raden är R multiplicerat med p , andra raden är R multiplicerat med $(1 - p)$ och sista raden är R multiplicerat med noll. Då fås att R och C är en uppdelning av matrisen T , så att $T = CR$ vilket betyder att $T_{ij} = C_i R_j$. Då fås K_L

$$K_L = -R\Sigma^{-1}C. \quad (3.38)$$

Genom att använda definitionen för K_L och matriserna i exempel 3.7 fås

$$\begin{aligned} K_L = -R\Sigma^{-1}C &= \begin{pmatrix} 0 & 0 & \beta \end{pmatrix} \Sigma^{-1} \begin{pmatrix} p \\ 1-p \\ 0 \end{pmatrix} \\ &= \left(\frac{p\beta v_1}{(v_1 + \mu)(\alpha + \mu)} + \frac{(1-p)\beta v_2}{(v_2 + \mu)(\alpha + \mu)} \right) \end{aligned} \quad (3.39)$$

På grund av att K_L är en endimensionell matris så är det samma som det domineranta egenvärdet, alltså får vi uttrycket för R_0 som nämnts tidigare.

Ett annat sätt som inte är lika vanligt att bestämma K_L är att bestämma en delmängd av infektionsstadie som är stadie av smittsamhet. Det här stadiet är då en person som är infekterad kan smitta en person som är mottaglig. Det här går att jämföra med att räkna nyfödda eller personer som blivit myndiga i en population. I exempel 3.7 finns det två stadier av infektion (E_1 och E_2) och ett stadie av smittsamhet (I). Då kan K minskas till att den endast visar stadiet av smittsamhet, beskriver generationsprocessen och det går att beräkna R_0 från K_L .

Sats 3.8. *Det gäller även allmänt att det största egenvärdet för K är också det största egenvärdet för både K_L och K_S .*

Bevis. Börjar med att visa att K och K_S har samma egenvärden som är olika noll. Låt $v \neq 0$ vara egenvektorn för K med egenvärdet λ . Då är $Kv = -\epsilon^\top T \Sigma^{-1} \epsilon v = \lambda v$. Multiplicera denna likhet med ϵ , och får $\epsilon(-\epsilon^\top) T \Sigma^{-1} \epsilon v = \epsilon \lambda v$. Från tidigare vet vi att $\epsilon \epsilon^\top T = T$, varför ϵv är egenvektorn för K_S med egenvärdet λ och då är egenvärden för K och K_L samma och olika noll. Det är omöjligt att $\epsilon v = 0$ eftersom det skulle implicera att $\lambda v = Kv = 0$ och därmed vore $v = 0$ för $\lambda \neq 0$.

Nu visas att K_S och K_L har samma egenvärden olika noll. Låt $v \neq 0$ vara egenvektorn för K_L med egenvärdet λ . Då är $K_L v = -R \Sigma^{-1} C v = \lambda v$. Multiplicera likheten med C för att få $-CR \Sigma^{-1} C v = \lambda C v$. Men $CR = T$, vilket betyder att Cv är egenvektorn till K_S med egenvärdet λ . Det är omöjligt att $Cv = 0$ för det skulle implicera att $\lambda v = K_L v = 0$ och därmed vore $v = 0$ för $\lambda \neq 0$.

Matriserna K , K_S och K_L har därmed samma egenvärden som är olika noll och därmed samma dominanta egenvärde och samma rank. \square

Sammanfattningsvis så har tre olika nästa generationsmatriser behandlats K , K_S och K_L . För att beräkna R_0 bör det största egenvärdet av K beräknas,

$R_0 = \rho(K)$. I de flesta fall är endast K och K_S relevanta och R_0 kan beräknas från någondera matriser. På grund av att K har färre dimensioner än K_S så är det enklare att komma till ett explicit uttryck för R_0 med hjälp av K . Ur epidemiologisk synvinkel beskriver elementen i matrisen $K = (k_{ij})$ stadiet av infektion och för varje stadie av infektion j har en person börjat sin infektion i stadie j . Följ sedan personen genom påföljande infektions liv och bestäm hur många nya fall hen är väntad att skapa för varje stadie av infektion i .

Kapitel 4

Första tillväxten av epidemin i realtid

Det här kapitlet kommer att handla om första tillväxten av epidemin i realtid och är främst baserat på boken *Mathematical Tools for understanding Infectious Disease Dynamics* [6]. Sannolikheten som beskriver att det inte nödvändigtvis blir en exponentiell tillväxt fastän $R_0 > 1$ betecknas med z_∞ . I enkla fall kan man härleda uttryck för R_0 och z_∞ .

Nackdelen med att mäta tillväxten i generationer är att de inte motsvarar det man observerar. Det här beror på att generationer ofta överlappar varandra. Därför kan epidemins exponentiella tillväxt i början skrivas som

$$Ce^{rt} \tag{4.1}$$

för något $r > 0$ som är antalet fall som registreras fram till tiden t och $C > 0$ är en konstant.

Definition 4.1. *Incidens, $i(t)$, är antalet nya fall per tidsenhet, till exempel per dag. Incidens är lika med graden av infektionen multiplicerat med delpopulationen som består av mottagliga personer.*

I det här fallet är incidensen proportionell till derivatan, e^{rt} . Incidensen är ofta lägre i insamlat data än vad den är i verkligheten eftersom alla fall inte rapporteras. Nya fall vid tiden t beror på antalet kontakter mottagliga personer har med personer som är smittade före tiden t . Det antas i en deterministisk beskrivning att antalet smittade personer är lika med det väntade värdet för

antalet smittade personer. När antaganden om kontakter, risken för smitta och elementära bokföringen läggs ihop fås ekvationen för incidensen i början av en epidemi

$$i(t) = pc \int_{T_1}^{T_2} i(t - \tau) d\tau \quad (4.2)$$

där p , T_1 , T_2 och τ beskrivs i (2.1). Ett nytt fall vid tiden t är resultatet av en kontakt med en infekterad som hade blivit infekterad tidigare, exempelvis vid tidpunkten $t - \tau$ för något positivt τ . För att vara smittsam bör τ vara i intervallet $T_1 \leq \tau \leq T_2$. Infekterade personer skapar kontakter med hastigheten c och smittar infektionssjukdomen vidare med sannolikheten p . Om vi delar upp incidensen, antalet nya fall per tidsenhet, i sannolikheterna för värdet av τ av smitta, alltså personen som är ansvarig för smitta, så fås (4.2).

Sats 4.2. *Substitutionen $i(t) = ke^{\lambda t}$, $k > 0$, $\lambda \in (-\infty, \infty)$ i (4.2) uppfyller ekvationen*

$$1 = pc \int_{T_1}^{T_2} e^{-\lambda\tau} d\tau =: f(\lambda). \quad (4.3)$$

Dessutom gäller $R_0 = pc(T_2 - T_1) = f(0) = 1$, $\lim_{\lambda \rightarrow \infty} f(\lambda) = 0$ och $\lim_{\lambda \rightarrow -\infty} f(\lambda) = \infty$.

Bevis. Insättning av $i(t) = ke^{\lambda t}$ i (4.2) ger

$$ke^{\lambda t} = kpc \int_{T_1}^{T_2} e^{\lambda t - \lambda\tau} d\tau, \quad (4.4)$$

och vidare emedan $k > 0$, $e^{\lambda t} > 0$, att

$$1 = pc \int_{T_1}^{T_2} e^{-\lambda\tau} d\tau. \quad (4.5)$$

För $\lambda > 0$ fås för högersidan av (4.3) att $1 = f(0) = pc(T_2 - T_1) = R_0$. Vidare gäller att

$$\frac{d}{d\lambda} \left(pc \int_{T_1}^{T_2} e^{-\lambda\tau} d\tau \right) = -pc \int_{T_1}^{T_2} \tau e^{-\lambda\tau} d\tau < 0, \quad (4.6)$$

varför $f(\lambda)$ är en avtagande funktion. Emedan $\lim_{\lambda \rightarrow \infty} \max_{T_1 \leq \tau \leq T_2} e^{-\lambda\tau} = 0$, följer att $\lim_{\lambda \rightarrow \infty} f(\lambda) = 0$. På samma sätt gäller att $\lim_{\lambda \rightarrow -\infty} \max_{T_1 \leq \tau \leq T_2} e^{-\lambda\tau} = \infty$, om $T_1 > 0$. Om $T_1 = 0$ så är gränserna för τ , $\gamma \leq \tau \leq T_2$ för fixerat $\gamma > 0$ och därmed gäller samma argument. I dessa fall fås $\lim_{\lambda \rightarrow -\infty} f(\lambda) = \infty$. \square

Det här kan sammanfattas som att det finns en reell rot $\lambda = r$, alltså visar (4.3) entydigt vad exponentiella ökningsgraden är för den här modellen. Det finns inte en explicit formel för r , i jämförelse med R_0 . Låt $r > 0$ om och endast om $R_0 > 1$ och $r < 0$ om och endast om $R_0 < 1$. Det här betyder att det endast finns en ökning i real tid om och endast om det finns en ökning i generationsmodellen. Notera även att $r = 0$ om och endast om $R_0 = 1$ och att det inte generellt gäller att för ett högt R_0 värde är r högt. Till exempel även om människor kan få många barn när de är relativt gamla så kan populationsökningen vara långsam. Ur ett matematiskt perspektiv så är R_0 endast beroende av $T_2 - T_1$, medan ekvationen för r har magnituden av T_1 och T_2 en stor betydelse. Därtill är R_0 dimensionslös medan r beror på enheten av tid som använts.

Kapitel 5

Kermacks och McKendricks modell

Det här kapitlet kommer främst att handla om Kermacks och McKendricks modell. Kapitlet börjar med en genomgång hur Kermack och McKendrick arbetade. Därefter går det igenom begrepp och definitioner, sedan matematiken kring modellen och tolkningar samt till sist en kort analys över artikeln [11] de publicerade år 1927 och dess popularitet.

5.1 Samarbete mellan Kermack och McKendrick

Det här avsnittet baserar sig på artikeln *The 1927 epidemic model of Kermack and McKendrick* [5].

Under 1800-talet fanns det en uppfattning om att dålig luftkvalitet orsakar malaria. Senare framkom det att malaria är ett virus men det var ännu oklart hur malaria smittar. År 1897 upptäckte Ronald Ross att malaria smittar via myggbett. Ross var mångbegåvad och fortsatte att undersöka malaria och kom fram till att om förhållandet mellan myggor och tätheten bland människor är tillräckligt låg kan man eliminera malaria.

År 1901 åkte Ross till Sierra Leone för att vidta åtgärder för att minska malaria fallen. Till arbetsgruppen hörde Anderson Grey McKendrick, 25 år, som hade studerat vid universitet i Glasgow, därefter genomfört militären och sedan studerat tropisk medicin. Under tiden i Sierra Leone väckte Ross intresse för matematik och spridning av infektionssjukdomar hos McKendrick, som visade sig vara begåvad i de här ämnena.

Efter tiden i Sierra Leone tillbringade McKendrick två årtionden i Indien där

han studerade sjukdomen rabies. Han återvände till Skottland 1920 och började arbeta på Royal College of Physicians i Edinburgh. Ett år senare började den 23 åriga biokemisten William Ogilvy Kermack arbeta på samma högskola och tre år senare blev han dessvärre blind på grund av en laboratorieexplosion. Trots olyckan slutade han inte att arbeta. Det är oklart när samarbetet mellan McKendrick och Kermack började men år 1927 gav de ut artikeln *A contribution to the mathematical theory of epidemics*. [11]

5.2 Beteckningar och definitioner

Det här avsnittet baserar sig främst på boken *Basic models in epidemiology* [3]. Kermacks och McKendricks modell är en grupperingsmodell där populationen är indelad i grupper. Modellen kallas *SIR* modellen och det finns tillämpningar *SIS*, eller *SEIR*, beroende på infektionssjukdomens beteende. Modellen baserar sig på att det finns en mottaglig population var det introduceras en obekant infektionssjukdom. Modellen fokuserar på att en liten grupp av personer blir infekterade som sedan smittar andra. Populationen är ofta delad i tre grupper *S*, *I* och *R*. Tid är en viktig variabel i modellen, som betecknas med t .

En annan modell är *SIS* modellen, där personer går från att varit infekterade till mottagliga, för det finns ingen immunitet för infektionssjukdomen. Sådana sjukdomar är vanligen orsakade av bakterier. Det finns även sjukdomar var personer inte lämnar infektionsstadiet, till exempel HIV. *SEIR* modellen är liknande som *SIR* men det finns en grupp *E* som betecknar gruppen som är exponerade.

Under en epidemi är populationsstorleken konstant. Det betyder att det inte sker en demografisk utveckling utan personer som dött under epidemin antas ha dött av sjukdomen. Det här antagandet är rimligt när det kommer till infektionssjukdomar som varar en kort tidsperiod, exempelvis en vecka. När en ny infektionssjukdom introduceras är alla personer i populationen mottagliga för smitta, populationen är med andra ord naiv. Det antas även att populationen är homogen. I *SIR* modellen antas det att personer tillfrisknar eller dör efter infektion [14].

- $S(t)$ är personer som är mottagliga vid tidpunkt t .
- $I(t)$ är personer som är infekterade och smittar personer som är mottagliga vid tidpunkten t .

- $R(t)$ är personer som inte är mottagliga för smitta och smittar inte andra vid tidpunkten t . Till gruppen hör personer som är immuna, karantän-satta, som har tillfrisknat och är därmed immuna eller dött på grund av sjukdomen.

Modellen skapas genom att derivera storleken av varje grupp och det antas att antalet personer i varje grupp är en differentierbar funktion över tid. Det här är fördelaktigt om det är många personer i varje grupp, annars kan resultatet bli snävt. Då fås Kermack och McKendricks modell [11],

$$\begin{cases} \frac{dS}{dt} = -\beta S \frac{I}{N} \\ \frac{dI}{dt} = \beta S \frac{I}{N} - \frac{1}{\tau} I \\ \frac{dR}{dt} = \frac{1}{\tau} I \end{cases} \quad (5.1)$$

där $N(t)$ är hela populationen som påverkas vid tidpunkten t och $N(t) = S(t) + I(t) + R(t)$. En person går från att vara i gruppen mottaglig, S , till gruppen infekterad, I , till gruppen död eller immun, R . Om gruppen $R(t)$ som har en egenskap att de inte mera kommer i kontakt med varken gruppen $S(t)$ eller $I(t)$ så är populationen $N(t) := S(t) + I(t)$.

Sannolikheten att en mottaglig person kommer i kontakt med en infekterad person är $\frac{I}{N}$. Kontaktprocessen för att sprida smitta är baserad på kontakter mellan mottagliga och infekterade. Låt $\beta = \beta(N)$ beteckna effektiva kontakthastigheten som är en funktion av storleken av den aktiva populationen. Det är vanligt att anta att $\beta = p\phi(N)$ där p är sannolikheten för smitta per kontakt och $\phi(N)$ är kontakthastigheten per person. Då fås att ϕS är medelvärdet av kontakter som mottagliga personer har per tidsenhet. Mottagliga personer har kontakter med personer från grupperna S , I och R , på grund av att det antas att gruppen är homogent blandat, så är kontakten med infekterade personer proportionerlig med $\frac{1}{N}$. Det totala medelvärdet för hur ofta kontakter sker är då $\phi S \frac{I}{N}$, dock är det viktigt att observera att alla kontakter inte leder till infektion. Den effektiva kontakthastigheten, även kallad incidens, är $\beta(N) S \frac{I}{N}$. Om N är konstant är ökningen av nya infektioner proportionerligt till både S och I . Det här gäller även om N är icke-konstant men $\beta(N) := pN$ det vill säga, en linjär funktion av N .

Att personer förflyttas från gruppen I till gruppen R är konstant $\frac{1}{\tau}$, per

person per tidsenhet. Det här antyder att fördelningen över hur lång tid en person tillbringar i grupp I är exponentiellt fördelad med medelvärdet τ .

Ett nyckelantagande i Kermack och McKendricks modell är incidensen av infektionssjukdom.

Definition 5.1. *Graden av infektion är sannolikheten per tidsenhet, till exempel en dag, att en mottaglig person blir infekterad.*

Om det antas att en person som blivit smittad överförs till gruppen infekterade och därmed kommer hen att påverka graden av infektion. Det här varar medan hen hör till gruppen infekterade och personen kommer att sluta påverka när hen tillfrisknar. Personen kommer att lämna gruppen med en konstant sannolikhet per tidsenhet. [5]

5.3 Matematiken i Kermacks och McKendricks modell

Det här avsnittet är främst baserat på artikeln *Basic models in epidemiology* [3] och på boken *Mathematical Tools for Understanding Infectious Disease Dynamics* [6].

På grund av att populationen är konstant så är $S + I + R = K$ eller som andelar av K , $\frac{1}{K}(S + I + R) = 1$. Det går att eliminera R från (5.1) genom att $R = K - S - I$. Vi kan bilda ett tvådimensionellt ekvationssystem,

$$\begin{cases} \frac{dS}{dt} = -\beta S \frac{I}{K} \\ \frac{dI}{dt} = \beta S \frac{I}{K} - \frac{1}{\tau} I \end{cases} \quad (5.2)$$

där β är antalet kontakter per tidsenhet och τ är mätt i tidsenheter.

I ekvationssystemet (5.2) är $\frac{dS(t)}{dt} < 0$ för alla t , det vill säga att S är avtagande då $t \geq 0$. För alla t är $\frac{dI(t)}{dt} < 0$ om och endast om $S < \frac{K}{\beta\tau}$. Om $S(0) < \frac{K}{\beta\tau}$ så är $S(t) \leq S(0) < \frac{K}{\beta\tau}$ som implicerar $\frac{dI(t)}{dt} < 0$ för alla t och infektionen dör ut. Men om $S(0) > \frac{K}{\beta\tau}$ så är $\frac{dI(t)}{dt} > 0$ vilket implicerar att $S(t) \geq \frac{K}{\beta\tau}$ på grund av att S är avtagande för t så minskar $S(t) < \frac{K}{\beta\tau}$ under epidemin och då är $\frac{dI}{dt} < 0$. Om få antal infekterade personer introduceras till en naiv population vid tidpunkten noll, så att $I(0) = \epsilon > 0$, $R(0) = 0$ och $S(0) = K - \epsilon \approx K$ fås tröskelvärdet

$$R_0 = \beta\tau. \quad (5.3)$$

Om $R_0 < 1$ dör infektionen ut och om $R_0 > 1$ bildas det en epidemi i populationen. För att förenkla analysprocessen antas det att $K = 1$. Det bör noteras att det här antagandet inte påverkar R_0 eftersom β antas vara konstant men det innebär en skillnad eftersom Kermack och McKendrick antog att $\beta = pK$. Alltså måste β mätas i andra enheter.

För att förenkla analysen av systemet (5.2) antas det att $K = 1$, då fås

$$\begin{cases} \frac{dS}{dt} = -\beta S(t)I(t) \\ \frac{dI}{dt} = \beta S(t)I(t) - \frac{1}{\tau}I(t) \end{cases} \quad (5.4)$$

Jämviktspunkterna för systemet ges då av,

$$\begin{cases} \frac{dS}{dt} = -\beta S(t)I(t) = 0 \\ \frac{dI}{dt} = \beta S(t)I(t) - \frac{1}{\tau}I(t) = 0. \end{cases} \quad (5.5)$$

En jämviktspunkt är $(S_\infty, 0)$, med godtyckligt S_∞ , då $0 \leq S_\infty \leq 1$, för $K = 1$.

Låt

$$\begin{cases} \frac{dS}{dt} = -\beta S(t)I(t) =: u(S, I) \\ \frac{dI}{dt} = \beta S(t)I(t) - \frac{1}{\tau}I(t) =: v(S, I) \end{cases} \quad (5.6)$$

Då fås som i avsnitt 1.2 att

$$\begin{pmatrix} \frac{dS}{dt} \\ \frac{dI}{dt} \end{pmatrix} = \begin{pmatrix} \frac{du}{dS}(S_\infty, 0) & \frac{du}{dI}(S_\infty, 0) \\ \frac{dv}{dS}(S_\infty, 0) & \frac{dv}{dI}(S_\infty, 0) \end{pmatrix} \begin{pmatrix} S \\ I \end{pmatrix} = \begin{pmatrix} 0 & -\beta S_\infty \\ 0 & \beta S_\infty - \frac{1}{\tau} \end{pmatrix} \begin{pmatrix} S \\ I \end{pmatrix}. \quad (5.7)$$

Låt $A = \begin{pmatrix} 0 & -\beta S_\infty \\ 0 & \beta S_\infty - \frac{1}{\tau} \end{pmatrix}$ och egenvärdena fås genom $\det(A - \lambda I_2) = 0$, där λ är egenvärden och I_2 är 2×2 enhetsmatrisen.

$$\begin{aligned} \det(A - \lambda I_2) &= \det \left(\begin{pmatrix} 0 & -\beta S_\infty \\ 0 & \beta S_\infty - \frac{1}{\tau} \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) \\ &= \det \begin{pmatrix} -\lambda & -\beta S_\infty \\ 0 & \beta S_\infty - \frac{1}{\tau} - \lambda \end{pmatrix} = -\lambda \left(\beta S_\infty - \frac{1}{\tau} - \lambda \right) = 0, \end{aligned} \quad (5.8)$$

varför $\lambda_1 = 0$ och $\lambda_2 = \beta S_\infty - \frac{1}{\tau}$. På grund av egenskaper hos egenvärden kan man inte dra slutsatser från egenvärdet $\lambda_1 = 0$. Däremot är egenvärdet $\lambda_2 = \beta S_\infty - \frac{1}{\tau}$ nyttigt för analysen. Emedan $|\lambda_2| > 0$, så är $R_0 = \max\{|\lambda_1|, |\lambda_2|\} = |\lambda_2|$

Gränsvärdet $S(\infty)$ beror på startpunkten $S(0)$, men $S(\infty) > 0$ för alla $S(0)$. Det betyder att alla personer i populationen inte kommer att bli infekterade.

Samma fenomen kan även observeras i data från tidigare epidemier. Nu påstås det att banan som går mellan punkterna $(S(0), 0)$ och $(S(\infty), 0)$ har sambandet,

$$S(\infty) - \frac{1}{\beta\tau} \ln S(\infty) = S(0) - \frac{1}{\beta\tau} \ln S(0). \quad (5.9)$$

Det fås från (5.4)

$$\frac{dI}{dS} = \frac{dI}{dt} \cdot \frac{dt}{dS} = \frac{\beta SI - \frac{1}{\tau} I}{-\beta SI} = -1 + \frac{1}{\tau\beta S}. \quad (5.10)$$

För att få I beräknas

$$\int \frac{dI}{dS} = \int -1 + \frac{1}{\tau\beta S} ds = -S + \frac{1}{\tau\beta} \ln S + C \quad (5.11)$$

var följer att

$$-S + \frac{1}{\tau\beta} \ln S + C = I \Leftrightarrow C = I + S - \frac{1}{\tau\beta} \ln S, \quad (5.12)$$

där C är en konstant. På grund av att $I(\infty) = I(0) = 0$ fås att

$$S(\infty) - \frac{1}{\beta\tau} \ln S(\infty) = S(0) - \frac{1}{\beta\tau} \ln S(0) = R_0. \quad (5.13)$$

I modellen antas det att kontakthastigheten β är konstant. Det här är inte alltid fallet utan β kan variera beroende på sjukdomens beteende och hur människor beter sig. Därför kan det vara svårt att uppskatta β och R_0 . Fastän β kan variera är $S(0) \approx K$ när en sjukdom introduceras i en population som är mottaglig och $S(\infty)$, andelen personer med antikroppar, det här kan uppskattas från serologiskt data. Den här relationen gör det möjligt att beräkna $\beta\tau$ när man har kunskap om $S(0)$ och $S(\infty)$. Då fås

$$\begin{aligned} S(0) - S(\infty) &= \frac{1}{\beta\tau} [\ln S(0) - \ln S(\infty)] \\ &= \frac{1}{\beta\tau} \ln \frac{S(0)}{S(\infty)} \Leftrightarrow \beta\tau = \frac{\ln \frac{S(0)}{S(\infty)}}{S(0) - S(\infty)}. \end{aligned} \quad (5.14)$$

Data i exemplet nedan är från artikeln *Viral infections of humans* [7].

Exempel 5.2. Under en epidemi vet man att $S(0) = 0,911$ och att $S(\infty) = 0,5133$ Då får från (5.14) att

$$R_0 = \beta\tau = \frac{\ln \frac{0,911}{0,5133}}{0,911 - 0,5133} = 1,44 > 1. \quad (5.15)$$

Det kan antas att τ är ungefär 3 dagar och då är $\beta \approx 0,48$ per dag. Så länge som $S(0) > \frac{1}{\beta\tau}$ är det möjligt att en epidemi bryter ut. I det här fallet har vi $0,911 > \frac{1}{1,44}$ det vill säga $0,911 > 0,694$, alltså det finns en möjlighet att en epidemi bryter ut.

5.3.1 SIS modellen

Den enklaste formen av *SIS* modellen är [11],

$$\begin{cases} \frac{dS(t)}{dt} = -\beta S(t) \frac{I(t)}{N} + \frac{1}{\tau} I(t) \\ \frac{dI(t)}{dt} = \beta S(t) \frac{I(t)}{N} - \frac{1}{\tau} I(t), \end{cases} \quad (5.16)$$

där $S + I = N$ och personer förflyttas tillbaka till grupp S med graden $\frac{1}{\tau}I$ istället för att förflyttas till grupp R . I det här fallet är populationen också konstant, $N = K$. Det kan antas som tidigare att $K = 1$ och förhållandet $I(t) = K - S(t)$ gäller för att förenkla modellen,

$$\frac{dS(t)}{dt} = \left(1 - \frac{S(t)}{K}\right) \left(-\beta S(t) + \frac{K}{\tau}\right) \quad (5.17)$$

med två jämviktspunkter

$$S(t) = K \text{ och } S(t) = \frac{K}{\beta\tau}. \quad (5.18)$$

Från den andra jämviktspunkten fås $S(t) = \frac{K}{\beta\tau}$ som är lika med $\frac{1}{\beta\tau} = \frac{S(t)}{K}$. Om $\frac{1}{\beta\tau} < \frac{S(t)}{K}$ eller $\beta\tau > 1$ så är båda jämviktspunkterna i intervallet $0 < S(t) \leq K$. Däremot, om $\frac{1}{\beta\tau} > \frac{S(t)}{K}$ eller $\beta\tau < 1$, är endast jämviktspunkten $S = K$ relevant. Det antas att β är konstant som implicerar att K kan vara lika med 1.

För att göra ekvationssystemet linjärt görs som i teorin i kapitel 1.2. Först linjäriseras systemet i jämviktspunkten $S(t) = 1$ (eller $I = 0$). Då fås

$$\frac{dS(t)}{dt} = (1 - S(t)) \left(-\beta S(t) + \frac{1}{\tau}\right) =: u(S) \quad (5.19)$$

och när $u(S)$ deriveras fås

$$\frac{du}{dS} = -\beta + 2\beta S - \frac{1}{\tau}. \quad (5.20)$$

När $S = 1$ fås att

$$\frac{du}{dS} = \beta - \frac{1}{\tau} \quad (5.21)$$

Jämvikten är asymptotiskt stabil om $\beta - \frac{1}{\tau} < 0$ eller $\beta\tau < 1$ och instabil om $\beta\tau > 1$.

Nu undersöks det linjäriserade ekvationssystemet i jämviktspunkten $S = \frac{1}{\beta\tau}$, som är intressant endast om $\beta\tau > 1$. Då fås

$$\frac{du}{dS} = \frac{1}{\tau} - \beta. \quad (5.22)$$

Jämvikten är asymptotiskt stabil om $\frac{1}{\tau} - \beta < 0$ eller $\beta\tau > 1$. Resultatet tyder på att det alltid finns en asymptotiskt stabil jämviktpunkt som lösningen tenderar mot. Om $\beta\tau < 1$ är jämviktpunkterna $S = 1$ och $I = 0$ tyder det på att infektionssjukdomen kommer att försvinna. Om $\beta\tau > 1$ och de asymptotiska stabila jämviktpunkterna är $S = \frac{1}{\beta\tau}$ och $I = 1 - \frac{1}{\beta\tau} > 0$ så kallas punkterna endemisk jämviktpunkt. Då kommer infektionssjukdomen att spridas.

5.3.2 SIER modellen

SIER modellen skiljer sig från *SIR* modellen genom att de finns en exponeringstid hos infektionssjukdomen. Med andra ord betyder det att en smittad person inte genast smittar andra utan först i infektionsstadiet I . Exponeringstiden är en bestämd längd, ω , med sannolikheten $Q(s)$ för att personer ska hållas kvar i gruppen exponerade vid tiden s efter att ha blivit exponerad för infektionssjukdomen. Analysen av infektionssjukdomsmodeller med exponerings- och infektionstid av bestämd längd minskar analysen av differential-differensekvationer med två förseningar i tid.

I exemplet nedan beaktas en *SEIR* modell, där E betecknar gruppen som är exponerad. *SIR* och *SIS* modellerna är speciella fall med exponeringstiden ω lika med 0.

Exempel 5.3. *Låt SEIR modellen vara utan födslar och dödlighet där $S + E + I + R = K$. Låt SEIR modellen ha en exponeringsperiod av bestämd längd ω och en infektionsperiod av bestämda längden τ . Då är*

$$\begin{cases} \frac{dS(t)}{dt} = -\beta S(t) \frac{I(t)}{K} \\ \frac{dE(t)}{dt} = \beta S(t) \frac{I(t)}{K} - \beta S(t - \omega) \frac{I(t - \omega)}{K} \\ \frac{dI(t)}{dt} = \beta S(t - \omega) \frac{I(t - \omega)}{K} - \beta S(t - \tau - \omega) \frac{I(t - \tau - \omega)}{K} \\ \frac{dR(t)}{dt} = \beta S(t - \tau - \omega) \frac{I(t - \tau - \omega)}{K} \end{cases} \quad (5.23)$$

och ekvationssystemet kan också skrivas som

$$\begin{cases} \frac{dS(t)}{dt} = -\beta S(t) \frac{I(t)}{K} \\ E(t) = \int_{t-\omega}^t \beta S(x) \frac{I(x)}{K} dx \\ I(t) = \int_{t-\tau-\omega}^{t-\omega} \beta S(x) \frac{I(x)}{K} dx \\ \frac{dR(t)}{dt} = \beta S(t - \tau - \omega) \frac{I(t - \tau - \omega)}{K}. \end{cases} \quad (5.24)$$

Vi kan skriva det här som två ekvationer, ty E och R är bestämda när S och I är bekanta, nämligen

$$\begin{cases} \frac{dS(t)}{dt} = -\beta S(t) \frac{I(t)}{K} \\ I(t) = \int_{t-\tau-\omega}^{t-\omega} \beta S(x) \frac{I(x)}{K} dx. \end{cases} \quad (5.25)$$

Det här kan skrivas som en ekvation,

$$I(t) = - \int_{t-\tau-\omega}^{t-\omega} \frac{dS(x)}{dx} dx = S(t - \tau - \omega) - S(t - \omega) \quad (5.26)$$

och då $\frac{dS(t)}{dt} = -\beta S(t) \frac{I(t)}{K}$ fås

$$\frac{dS(t)}{dt} = -\beta S(t) \frac{[S(t - \tau - \omega) - S(t - \omega)]}{K}. \quad (5.27)$$

För att den här enkla differential-differensekvationen med två förseningar ska gälla måste det läggas till data från starten för att $\tau + \omega \leq t \leq 0$ ska vara ett välformulerat problem. Det här beror på att varje konstant är en lösning till differential-differensekvationen. Det är möjligt att bevisa att $S(t)$ är avtagande och närmar sig en gräns $S(\infty) > 0$ då $t \rightarrow \infty$. Gränsvärdet kommer att bero på ursprungsdata. Analysen av den här modellen är mer komplicerad än analysen av Kermacks och McKendricks ursprungliga SIR modell, men tröskelvärdet för reproduktionstalet är samma för båda modellerna.

5.3.3 Exempel av SIR modellen

Mässling är en sjukdom som har undersökts i flera olika populationer med olika jämviktspunkter. Soper [15] formulerade en *SIR* modell år 1929 med en konstant grad av födslar μK i gruppen S och en konstant takt av dödlighet μK i gruppen R . Då fås

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta S}{K} I + \mu K \\ \frac{dI}{dt} = \frac{\beta S}{K} I - \frac{1}{\tau} I \\ \frac{dR}{dt} = \frac{1}{\tau} I - \mu K. \end{cases} \quad (5.28)$$

Den här modellen är inte realistisk eftersom kopplingen mellan födsel bland mottagliga och dödlighet är ohållbar. Den är också matematiskt ohållbar om $R(0)$ och $I(0)$ är tillräckligt små eftersom då kommer $R(t)$ att bli negativt. Den här modellen antar att β är en funktion av K och då bör man vara noggrann med enheterna.

Hethcote [9] kom fram till att bristerna i Sopers modell löses med att anta att dödlighet i respektive grupp är proportionell med antalet personer i gruppen. Det leder till att populationsstorleken hålls konstant. Nu fås följande modell,

$$\begin{cases} \frac{dS}{dt} = -\beta S \frac{I}{K} + \frac{1}{L}(K - S) \\ \frac{dI}{dt} = \beta S \frac{I}{K} - \frac{1}{\tau}I - \frac{1}{L}I \\ \frac{dR}{dt} = \frac{1}{\tau}I - \frac{1}{L}R, \end{cases} \quad (5.29)$$

där L är medelvärdet för livslängden då $\frac{1}{L}$ är graden för dödlighet. Här gäller även att $S + I + R = K$. Då kan ekvationssystemet skrivas om genom att eliminera R , och det följer

$$\begin{cases} \frac{dS}{dt} = -\beta S \frac{I}{K} + \frac{1}{L}(K - S) \\ \frac{dI}{dt} = \beta S \frac{I}{K} - \frac{1}{\tau}I - \frac{1}{L}I = I \left(\beta \frac{S}{K} - \frac{1}{\tau} - \frac{1}{L} \right). \end{cases} \quad (5.30)$$

Jämviktspunkterna fås genom att

$$\begin{cases} \frac{dS}{dt} = -\beta S \frac{I}{K} + \frac{1}{L}(K - S) = 0 \\ \frac{dI}{dt} = \beta S \frac{I}{K} - \frac{1}{\tau}I - \frac{1}{L}I = 0 \end{cases} \quad (5.31)$$

eller

$$\begin{cases} \beta \frac{SI}{K} = \frac{1}{L}(K - S) \\ I \left(\beta \frac{S}{K} - \frac{1}{\tau} - \frac{1}{L} \right) = 0. \end{cases} \quad (5.32)$$

En jämviktspunkt är $S(t) = K$, $I = 0$ (infektionssjukdom försvinner). Om $\frac{dI}{dt} > 0$, alltså om $\left(\beta \frac{S}{K} - \frac{1}{\tau} - \frac{1}{L} \right) > 0$,

$$\beta \frac{S(t)}{K} - \frac{1}{\tau} - \frac{1}{L} > 0 \Leftrightarrow \beta \frac{S(t)}{K} > \frac{1}{\tau} + \frac{1}{L} \Leftrightarrow \frac{\tau + L}{\beta \tau L} < \frac{S(t)}{K}. \quad (5.33)$$

I början av epidemin är $S(0) \approx K$ och då fås

$$\frac{\tau + L}{\beta \tau L} < 1 \Leftrightarrow \beta \tau \left(\frac{L}{L + \tau} \right) > 1. \quad (5.34)$$

Det finns en annan jämviktspunkt $S(t) = \frac{\tau + L}{\beta \tau L} K$ med $I > 0$ (endemisk jämviktspunkt). Genom linjärisering är det enkelt att visa om

$$\beta \tau \left(\frac{L}{L + \tau} \right) < 1, \quad (5.35)$$

är jämviktspunkten $S = K$, $I = 0$ asymptotiskt stabil, medan om

$$\beta\tau \left(\frac{L}{L + \tau} \right) > 1 \quad (5.36)$$

är jämviktspunkten instabil.

Tröskelvärdet

$$R_0 = \beta\tau \left(\frac{L}{L + \tau} \right) \quad (5.37)$$

är analogt med Kermack och McKendricks modell förutom att τ måste korrigeras för att ta i beaktande naturlig dödlighet under infektionsperioden, alltså korrigerar

$$\tau \left(\frac{L}{L + \tau} \right) \quad (5.38)$$

medeltalet över dödlighet under infektionsperioden.

Jämviktspunkten $S = K$, $I = 0$ är alltså instabil och man kan visa att endemiska jämviktspunkten är asymptotiskt stabil om $R_0 > 1$. Således tillåter modellen endemiska jämviktspunkter. Både SIR modellen med, födsel och dödlighet, och SIS modellen stöder endemiska jämviktspunkter. Alltså är kravet för stödet av endemiska jämviktspunkter flödet av nya mottagliga, antingen genom födslar eller tillfrisknande utan immunitet.

Generellt om $\beta(N) := p\phi(N)$ är villkoret $R_0 > 1$ ekvivalent med

$$\phi(S(0)) > \frac{\tau + L}{p\tau L}, \quad (5.39)$$

eller om ϕ har en invers med villkoret

$$S(0) > N_{kritisk} = \phi^{-1} \left(\frac{\tau + L}{p\tau L} \right). \quad (5.40)$$

Alltså måste den mottagliga populationen nå ett kritiskt värde för att en epidemiska uppstå. Uppskattningen av det kritiska värdet N beror på kunskapen av funktionen $\beta(N)$. Grova uppskattningar av $N_{kritisk}$ har gjorts på basen av antagandet $\beta(N) = pN$. Då har $N_{kritiskt}$ beräknats från indirekt data med formeln

$$N_{kritiskt} = \frac{\tau + L}{p\tau L}. \quad (5.41)$$

Ett exempel med de här uppskattningarna är att åtminstone 300 000 personer kommer att insjukna i influensa i städer på fastlandet medan åtminstone 500 000 personer kommer att insjukna på öar. Orsaken att värdena är olika är för att p och

infektionsåldern av infektionen varierar. Ett annat alternativ är att använda data. Då fås, om N är minst 300 000 i städer och p är tillräckligt stort, att endemin uppehålls. Samma resultat gäller för öar så länge som N är minst 500 000. För att summera dessa resultat måste data vara konsistent med samma funktion $\beta(N)$, alltså då måste p variera mellan öar och städer. [1]

Tröskelvärden är viktiga för att bestämma tillvägagångssättet för att kontrollera och utrota smittspridningen av sjukdomen. För att en infektionssjukdom ska utrotas bör $R_0 < 1$. Det här kan ske genom att personer i populationen blir immuna efter insjuknande och då flyttas personen från grupp S till R och då minskar K . Graden av immuniseringen av en andel, q , av populationen är samma som att byta ut K till $K(1 - q)$. Därmed ersätts reproduktionstalet $R_0 = \beta\tau \left(\frac{L}{L+\tau}\right)$ med $\beta\tau \left(\frac{L}{L+\tau}\right) (1 - q)$. Villkoret

$$\beta\tau \left(\frac{L}{L+\tau}\right) (1 - q) < 1 \text{ ger } 1 - q < \frac{1}{R_0} \text{ eller } q > 1 - \frac{1}{R_0}, \quad (5.42)$$

som garanterar att det lyckats.

Flockimmunitet bildas när en tillräckligt stor andel är immuna så att infektionssjukdomen inte sprids mera. Exempel på en sjukdom som hela världen är immun mot är smittkoppor. Det har man lyckats uppnå genom vaccinering.

5.3.4 Smittsamhetstakt och väntetid

I de tidigare modellerna har det antagits att takten av förflyttning mellan grupperna har varit proportionerlig till storleken av gruppen, som till exempel graden av tillfrisknande $\frac{1}{\tau}I$. I verkligheten betyder inte dessa antaganden att tillfrisknande beror på storleken av gruppen infekterade. En bättre formulering i stället för differentialekvationen

$$\frac{dI(t)}{dt} = \beta S(t) \frac{I(t)}{N} - \frac{1}{\tau} I(t) \quad (5.43)$$

är

$$I(t) = \int_0^t \beta S \frac{I(x)}{N(x)} e^{-\frac{1}{\tau}(t-x)} dx. \quad (5.44)$$

Om t är tillräckligt stort för att personer som var infekterade i början då $t = 0$ som har återhämtat sig och vi antar att $N(x) = K$. Då är integralekvationen

ekvivalent med differentialekvationen. Då ger derivering av integralen

$$\begin{aligned} \frac{dI(t)}{dt} &= \beta S(t) \frac{I(t)}{K} + \int_0^t \beta S(x) \frac{I(x)}{K} \frac{d}{dt} e^{-\frac{1}{\tau}(t-x)} dx \\ &= \beta S(t) \frac{I(t)}{K} - \frac{1}{\tau} \int_0^t \beta S(x) \frac{I(x)}{K} e^{-\frac{1}{\tau}(t-x)} dx \\ &= \beta S(t) \frac{I(t)}{K} - \frac{1}{\tau} I(t). \end{aligned} \quad (5.45)$$

Integralekvationen tolkas så att $e^{-\frac{1}{\tau}s}$ är sannolikheten att man är infekterad vid tidpunkten s efter att man blivit infekterad. Låt

$$\beta S(x) \frac{I(x)}{K} e^{-\frac{1}{\tau}(t-x)} \quad (5.46)$$

beskriva antalet personer som blir infekterade vid tidpunkt x och är infekterade $(t-x)$ tidsenheter senare. Integralen

$$\int_0^t \beta S(x) \frac{I(x)}{K} e^{-\frac{1}{\tau}(t-x)} dx \quad (5.47)$$

är integralen över tiden $x \leq t$ av antalet infekterade vid tidpunkt t . Genomsnittliga infektionstiden är

$$\int_0^\infty e^{-\frac{1}{\tau}s} ds = \tau. \quad (5.48)$$

En liknande tolkning kan anpassas för proportionalitetskonstanten $\frac{1}{L}$ för dödlighetsgraden i grupperna i *SIR* modellen var födsel och dödlighet beaktas. Låt

$$e^{-\frac{1}{L}s} \quad (5.49)$$

vara sannolikheten för överlevnad till åldern s och därmed är L förväntade genomsnittliga livslängden.

Det går att ge en tolkning för den väntade tiden av graden av infektion i *SIR* modellen med födsel och dödlighet. I den endemiska jämviktspunkten (S_∞, I_∞) är graden av nya infektioner $\beta S_\infty \frac{I_\infty}{K}$. Låt $\beta \frac{I_\infty}{K}$ vara graden av infektion per mottaglig person. Om det antas att en exponentiell sannolikhetsfördelning över infektion av en person, beskriver medelvärdet för tiden en person tillbringar i gruppen mottagliga före infektion är $\frac{K}{\beta I_\infty}$. Låt då medelvärdet för infektionsåldern vara $A = \frac{K}{\beta I_\infty}$. Låt (S_∞, I_∞) vara den endemiska jämviktspunkten som tidigare, med $I_\infty > 0$ som löser ekvationerna

$$\begin{cases} \beta S_\infty \frac{I_\infty}{K} = \frac{1}{L}(K - S_\infty) \\ \frac{I_\infty}{K}(\beta S_\infty - \frac{1}{\tau} - \frac{1}{L}) = 0 \end{cases} \quad (5.50)$$

varav det fås

$$\begin{cases} S_\infty = \frac{\tau+L}{\beta\tau L}K \\ I_\infty = \frac{K}{\beta LS_\infty}(K - S_\infty) \end{cases} \quad (5.51)$$

Reproduktionstalet är då

$$R_0 = \beta\tau \left(\frac{L}{\tau + L} \right) = \frac{K}{S_\infty}, \quad (5.52)$$

med $R_0 > 1$. Från dessa förhållanden fås att

$$\frac{L}{A} = \beta L \frac{I_\infty}{K} = \frac{K}{S_\infty} \left(1 - \frac{S_\infty}{K} \right) = \frac{K}{S_\infty} - 1 = R_0 - 1. \quad (5.53)$$

Relationen

$$R_0 = 1 + \frac{L}{A} \quad (5.54)$$

är användbar för att uppskatta reproduktionstalet då L och A kan mätas. Notera att denna härledning antar att infektionsstadiet är endemiskt och därmed är $R_0 > 1$.

Ett specialfall är att i stället för att anta att graden av tillfrisknande är exponentiellt fördelad, anta att $P(s)$ är sannolikheten för återstående infektion vid tidpunkten s efter att personen blivit infekterad. $P(s)$ är en avtagande funktion med $P(0) = 1$. Låt

$$\tau = \int_0^\infty P(s) ds \quad (5.55)$$

vara genomsnittet av infektionstiden. Ett viktigt specialfall är

$$P(s) = \begin{cases} 1, & 0 \leq s \leq \tau \\ 0, & s > \tau \end{cases} \quad (5.56)$$

motsvarande till en bestämd infektionsperiod τ . Då är integralekvationen för $I(t)$

$$I(t) = \int_0^t \beta S(x) I(x) P(t-x) dx \quad (5.57)$$

som kan skrivas som

$$I(t) = \int_{t-\tau}^t \beta S(x) I(x) dx \quad (5.58)$$

som är ekvivalent med differentialekvationen

$$\frac{dI(t)}{dt} = \beta S(t) I(t) - \beta S(t-\tau) I(t-\tau). \quad (5.59)$$

5.4 Kermack och McKendricks artikels framgång

I början av september 2021 har Kermacks och McKendricks artikel [11] citerats 103 330 gånger och artikeln har laddats ner från Royal Society 49 938 (jämfört med 2019, då hade artikeln citerats 6 230 gånger och laddats ner 21 695 gånger) sen den publicerades elektroniskt år 1997. Det här kan tolkas som att artikeln är lyckad. Enligt Diekmann finns det orsak till tvivel att artikeln läses ordentligt. Orsaken är att det finns en tro om att artikeln handlar om *SIR* systemet, vilket kan bero på att den inte lästs utan personer har hört eller läst om den i artiklar eller andra sammanhang. En annan orsak är att *SIR* modellen (5.1) är en standardform av att skapa modeller för tillämpade matematiker och att många teorier och verktyg finns. Däremot har förnybara ekvationer oändligt många dimensioner. Enligt Diekmann är artikeln, [11] en bra artikel men det är vanligt att de viktiga insikterna förbises. [5]

Kapitel 6

Den slutliga storleken på en epidemi

I en sluten population kommer andelen av mottagliga personer att minska när infekterade personer tillfrisknar eller dör. Det här betyder att en epidemi är ändlig i tid som går mot oändligheten. Det här kapitlet är baserat på boken *Mathematical Tools for understanding Infectious Disease Dynamics* [6].

Låt $s = S/N$ vara andelen personer som är mottagliga för smitta i en sluten population N . Låt $s(\infty)$ vara andelen mottagliga personer i slutet av epidemins utbrott. Komplementet för $s(\infty)$ är $1 - s(\infty)$ som är andelen av populationen som blivit infekterade, alltså slutliga storleken av epidemin. Ett viktigt resultat är att $s(\infty)$ är en lösning av den *slutliga storleksekvationen*

$$\ln s(\infty) = R_0(s(\infty) - 1) \quad (6.1)$$

och kan även skrivas som

$$s(\infty) = e^{-R_0(1-s(\infty))}. \quad (6.2)$$

Ekvation (6.1) kommer från (5.12),

$$\frac{1}{\beta\tau} \ln S(t) - S(t) - I(t) = \frac{1}{\beta\tau} \ln S(0) - S(0) - I(0) \quad (6.3)$$

Med $s(t) = S(t)/N$ fås att

$$\ln s(t) = \ln s(0) + N\beta\tau \left[s(t) - \frac{S(0) + I(0)}{N} + \frac{I(t)}{N} \right]. \quad (6.4)$$

Då $t \rightarrow \infty$ går $I(t) \rightarrow 0$. Låt $S(0) + I(0) = N$. Då $\lim_{t \rightarrow \infty} s(t) = 1$ så fås således då $t \rightarrow \infty$ att

$$\ln s(\infty) = R_0(s(\infty) - 1), \quad (6.5)$$

där $R_0 = N\beta\tau$.

Sats 6.1. *Då $R_0 < 1$ är den relevanta lösningen $s(\infty)$ till (6.1) given av $s(\infty) = 1$. Då $R_0 > 1$ finns den relevanta lösningen $s(\infty)$ i intervallet $(0, 1)$.*

Bevis. Funktionen $h(s) = \ln s - R_0(s - 1)$ uppfyller $h(1) = 0, h'(1) = 1 - R_0$, $h''(s) = -1/s^2 < 0$ för $0 < s < 1$. Därtill, $h(s) \downarrow -\infty$ då $s \downarrow 0$ och $h(s) \downarrow -\infty$ då $s \rightarrow \infty$. Egenskapen $h''(s) < 0$ garanterar att det finns högst två rötter. För att $h(1) = 0$ och $h'(1) = 1 - R_0$, vet vi att $h(s) > 0$ för $s < 1$ då $R_0 < 1$ och för $s > 1$ då $R_0 > 1$. Kontinuiteten för h och faktumet att $h(s) \downarrow -\infty$ för både $s \downarrow 0$ och $s \rightarrow \infty$ implicerar att för $R_0 \neq 1$ finns det alltid ett annat nollställe som är större än ett och därmed irrelevant då $R_0 < 1$. Då nollställe är mindre än ett och kallas $s(\infty)$ då $R_0 > 1$. För specialfallet $R_0 = 1$ har vi $s = 1$ som är ett dubbel nollställe. När $R_0 < 1$ fås att $s(\infty) = 1$ vilket betyder att den nya infektionen som introducerats i den slutna populationen inte leder till ett omfattande utbrott av en epidemi. När $R_0 > 1$ existerar det en entydig lösning i intervallet $(0, 1)$. \square

Det går att dra slutsatsen att en viss andel av $s(\infty)$ slipper att bli infekterade som beror på R_0 i ekvation (6.2). Ju större R_0 är, desto mindre andel av populationen blir inte infekterade. Andelen blir egentligen obefintligt liten om R_0 är stort.

6.1 Den slutliga storleken av en epidemi i en population utan restriktioner

Det antas att det för en population inte vidtas åtgärder för att minska smittspridningen. Det antas även att infektionen i fråga är såpass allvarlig att det finns en stor sannolikhet att den leder till död. När infektionen sprider sig i populationen minskar populationen men infektionens spridning fortsätter i samma grad. I det här fallet är antalet kontakter per tidsenhet oberoende av populationsstorleken.

Det här leder till att graden av epidemin inte minskar så snabbt som det tidigare antogs då populationstätheten och antalet kontakter mellan personer per tidsenhet är proportionell till populationens storlek. Det finns två parametrar i den slutliga storleksekvationen, sannolikheten att man överlever infektionen f och R_0 . Populationen kommer att ha två delar: andelen som inte får sjukdomen som betecknas med $s(\infty)$ och andelen som överlevt infektionen som betecknas med $n(\infty)$. Låt $n(\infty)$, $s(\infty)$ och konstanten f ha relationen

$$n(\infty) - s(\infty) = f(1 - s(\infty)). \quad (6.6)$$

För att komma fram till ett annat samband mellan $n(\infty)$, $s(\infty)$ och R_0 ser vi på följande exempel.

Exempel 6.2. *Beakta följande ekvationssystem,*

$$\begin{cases} \frac{dS}{dt} = -\gamma \frac{SI}{N} \\ \frac{dI}{dt} = \gamma \frac{SI}{N} - \alpha I \\ \frac{dN}{dt} = -(1-f)\alpha I, \end{cases} \quad (6.7)$$

där γ är produkten av väntade antalet kontakter per tidsenhet och sannolikheten för att smitta ska ske. Längden av infektionstiden är exponentiellt fördelad med parametern α . Om man beräknar, så fås

$$\frac{dS}{dN} = \frac{\gamma \frac{SI}{N}}{(1-f)\alpha I} = \frac{\gamma S}{(1-f)\alpha N} = \frac{R_0 S}{(1-f)N} \quad (6.8)$$

där $R_0 = \frac{\gamma}{\alpha}$ i det här fallet. Vi omskriver

$$\frac{dS}{dN} = \frac{R_0 S}{(1-f)N} \quad (6.9)$$

till

$$\frac{dS}{S} = \frac{R_0}{1-f} \frac{dN}{N} \quad (6.10)$$

och integrerar båda sidorna från σ till t . Då fås

$$\ln \frac{S(t)}{S(\sigma)} = \frac{R_0}{1-f} \ln \frac{N(t)}{N(\sigma)}. \quad (6.11)$$

Nu följer att

$$\frac{S(t)}{S(\sigma)} = \left(\frac{N(t)}{N(\sigma)} \right)^{\frac{R_0}{1-f}}. \quad (6.12)$$

Då $\sigma \downarrow -\infty$ går både $S(\sigma)$ och $N(\sigma)$ mot ursprungliga storleken av populationen. Så när vi beräknar gränsvärdena då $\sigma \downarrow -\infty$ och $t \uparrow \infty$ fås sambandet

$$n(\infty) = s(\infty)^{\frac{(1-f)}{R_0}}. \quad (6.13)$$

Andelen som varit infekterade som överlevt är $n(\infty) - s(\infty)$ och andelen infekterade är $1 - s(\infty)$. Om f betecknar sannolikheten för överlevnad gäller (6.6). Om f och R_0 är okända men det finns data efter epidemin som gör det möjligt att estimeras $s(\infty)$ och $n(\infty)$ kan den slutliga storleksekvationen uppskattas,

$$f = \frac{n(\infty) - s(\infty)}{1 - s(\infty)}, \quad R_0 = \frac{1 - n(\infty)}{1 - s(\infty)} \frac{\ln s(\infty)}{\ln n(\infty)}. \quad (6.14)$$

Man får $f = \frac{n(\infty) - s(\infty)}{1 - s(\infty)}$ direkt från (6.6). Däremot fås $R_0 = \frac{1 - n(\infty)}{1 - s(\infty)} \frac{\ln s(\infty)}{\ln n(\infty)}$ genom att räkna logaritmen av (6.13), ty

$$\begin{aligned} R_0 &= (1 - f) \frac{\ln s(\infty)}{\ln n(\infty)} \\ &= \frac{1 - s(\infty) - n(\infty) + s(\infty)}{1 - s(\infty)} \frac{\ln s(\infty)}{\ln n(\infty)} \\ &= \frac{1 - n(\infty)}{1 - s(\infty)} \frac{\ln s(\infty)}{\ln n(\infty)}. \end{aligned} \quad (6.15)$$

När personer dör finns det en direkt och en indirekt effekt. Den direkta effekten är att andelen $1 - f$ av $1 - s(\infty)$ dör. Den indirekta effekten är att $s(\infty)$ avtar eftersom immuna personer hindrar smittspridning, däremot kan inte döda personer göra det. Den indirekta effekten är det som gör att modeller med populationer som vidtar åtgärder och populationer som inte vidtar åtgärder är olika. När R_0 är stort och $s(\infty)$ är litet (ungefär e^{-R_0}) är indirekta effekten obetydlig. Däremot om R_0 är något ovanför tröskelvärdet kan den indirekta effekten vara betydande.

6.2 Den slutliga storleken av en epidemi i en ändlig population

I den deterministiska beskrivningen av en epidemi i en sluten population fanns beteckningen $1 - s(\infty)$ som beskrev andelen som blir påverkade av epidemin. Det antas att en population består av N personer och av dessa personer har en

person blivit smittad från en utomstående källa. När delmodellerna för kontakter och smitta har specificerats kan sannolikhetsfördelningen för antalet nya infektioner bestämmas. I den här delen kommer binomialfördelningen att användas i stället för poissonfördelningen, eftersom det är ändligt många personer som är mottagliga för smitta i populationen.

När $R_0 < 1$ är fördelningen för slutliga storleken av epidemin nära 0. När $\frac{k}{N}$ används i stället för k , som beskriver den slutliga storleken, och då $N \rightarrow \infty$, närmar sig fördelningen 0. Det betyder att den obetydliga andelen av populationen påverkas.

När $R_0 > 1$ har fördelningen för den slutliga storleken av epidemin en dubbel topp. Den första toppen beskriver det mindre utbrottet var infektionen dör ut innan den hinner påverka en stor andel av populationen. Den andra toppen beskriver det större utbrottet, var ungefär $1 - s(\infty)$ av populationen blir smittade. Delningen mellan den mindre och större utbrottet har en större betydelse om $\frac{k}{N}$ skalas och $N \rightarrow \infty$. Då är andelen z_∞ av totala sannolikheten, 1, centrerad kring 0, medan $1 - z_\infty$ närmar sig $1 - s(\infty)$. Om $1 - s(\infty)$ skalas lämpligt beskrivs den av normalfördelningen med standardavvikelsen $1/\sqrt{N}$. Det finns två observationer gällande det deterministiska framställningssättet. Även om $R_0 > 1$ kan det bildas endast ett mindre utbrott och $1 - s(\infty)$ är medelvärdet för storleken av det större utbrottet. Det har introducerats flera olika sätt att klassificera olika infektionssjukdomar beroende på populationstätheten, kontaktprocessen och sannolikheten att infektionen smittar. Det finns ett samband mellan de olika delarna

$$R_0 > 1 \iff r > 0 \iff s(\infty) < 1 \iff z_\infty < 1, \quad (6.16)$$

r definieras i kapitel 4.

När modeller skapas ska man beakta om latensperioden är lång för infektionssjukdomen. Om latensperioden är lång kan det ha en stor påverkan på R_0 men liten verkan på r . Däremot är ordningen för R_0 och $s(\infty)$ för två icke-dödliga infektioner alltid lika om det är frågan om likformigt fördelade grupper.

Kapitel 7

Modellering av COVID-19

Det här kapitlet baserar sig på artikeln *A SIR-type model describing the successive waves of COVID-19* [13]. Som tidigare nämnts kan *SIR* modellen inte göra precisa uppskattningar när det gäller infektionssjukdomar som COVID-19. I det här avsnittet har *SIR* modellen anpassats så att det går att göra lämpliga uppskattningar med officiella data från Italien, Spanien och USA gällande ökningen av COVID-19. I den här anpassningen av *SIR* modellen antas det att populationen inte är konstant och graden av infektion och dödlighet minskar exponentiellt. På så vis får man även information om antalet aktiva fall, kumulativ dödlighet, dagliga antalet nya fall och dödsfall som är orsakade av COVID-19. Man har provat modellen och den passar till existerande data från den första successiva vågen av COVID-19 i Italien, Spanien och USA. Orsaken att man valt de här länderna är att de har påverkats allvarligt av pandemin och har publicerat data kring COVID-19. Det kommer även att presenteras uppskattningar av hur pandemin kommer att fortskrida i Italien och USA med flera olika scenarier.

Ett problem med *SIR* modellen är att det antas att populationen är konstant, vilket kan stämma relativt bra när det är fråga om en infektionssjukdom som inte varar längre än några veckor. Däremot när det är fråga om COVID-19 så finns det en demografisk utveckling då pandemin bröt ut under vintern 2020 och är ännu aktiv under hösten 2021. Det stämmer inte överens med verkligheten att graden av smitta som sker vid kontakt och att graden av personer som flyttar från gruppen infekterade till gruppen immuna är konstant. När de här parametrar är konstanta beaktas inte åtgärder som vidtagits för att minska smittspridningen. I artikeln [13] har man använt en typ av autonom *SIR* modell som beaktar att

parametrarna inte är konstanta. När modellen appliceras på de här tre länderna kalibreras den enligt data som samlats under en tidsperiod som är minst 30 dagar. Man beaktar data som samlats in minst tio dagar efter att det har utlysts begränsningar i respektive land.

Den anpassade *SIR* modellen ser ut på följande sätt

$$\begin{cases} \frac{dS}{dt} = (\lambda - \mu)S + \lambda(I + R) - \beta SI \\ \frac{dI}{dt} = \beta SI - \gamma I - \frac{d\mu}{dt}I - \mu I \\ \frac{dR}{dt} = \gamma I - \mu R, \end{cases} \quad (7.1)$$

där graden av smittsamhet β och graden av dödlighet $\frac{d\mu}{dt}$ orsakad av COVID-19 inte är konstanta. Från data kan man se att i områden där man i början av pandemin har applicerat strikta begränsningar gällande att hålla socialt avstånd, har både β och $\frac{d\mu}{dt}$ minskat exponentiellt jämfört med deras standardvärden β_0 och $\frac{d\mu_0}{dt}$. Det har troligtvis att göra med att utvecklingen av variablerna är flyktig och därför behövs det tid för att de ska nå deras asymptotiska värde. Tillfriskningsgraden γ är inte nödvändigtvis konstant. Dock har observationer utifrån data visat att γ är relativt stabil i de tre länderna. Månatliga förändringar i γ skapar betydande förändringar i de andra epidemiologiska parametrarna. Graden av födsel $\lambda > 0$ är konstant och samma för alla tre kategorier S , I och R . Det antas även att alla avkommor är mottagliga för infektionen och att de inte är immuna när de föds. Det antas även att dödlighetsgraden $\mu > 0$ då det förekommer dödlighet som är orsakad av annat än COVID-19. Värden för λ och μ fås från tabell 7.1.

Land	Graden av födsel λ	Graden av dödlighet μ	Population N
Italien	$1,99258 \cdot 10^{-5}$	$2,866523 \cdot 10^{-5}$	$6,05 \cdot 10^7$
Spanien	$2,174293 \cdot 10^{-5}$	$2,494899 \cdot 10^{-5}$	$4,7 \cdot 10^7$
USA	$3,178082 \cdot 10^{-5}$	$2,356164 \cdot 10^{-5}$	$3,282 \cdot 10^8$

Tabell 7.1: Tabell över värden för λ och μ år 2018 [13].

Det ska inte tas för givet att tillfriskningsgraden γ är konstant, även om data tyder på det. Tillfrisknande eller dödligheten beror på hur personen som insjuknat reagerar på viruset. Det kan väntas att i ett övergående stadie finns variationer på graden av tillfrisknande och dödlighet tills asymptotiska värden är

tillgängliga. Det här verkar ha skett i USA och Italien gällande tillfrisknande och dödlighet. Oberoende antas det att γ är konstant under varje kalendermånad eller under undersökningsperioden i varje land. Låt γ vara baserat på medelvärdet av observerade månatliga värdet för γ . Normalt är γ likformigt fördelat med få undantag. För att bestämma β och $\frac{d\mu}{dt}$ har det observerats en signifikant minskning i infektions- och dödlighetsgraden i data under studien. Funktionerna β och $\frac{d\mu}{dt}$ är definierade genom en exponentiell regression. Enligt data närmar sig dödlighetsgraden sitt asymptotiska värde under ett relativt kort tidsintervall. Det går att förklara med att det sker förändringar i sociala beteenden som kan skapa en övergångsfas där parametrarna i modellen kan ändra och sedan närmar de sig sitt standardvärde.

I undersökning används en normaliserad modell,

$$\begin{aligned} S_N(t) &= \frac{S(t)}{N} \\ I_N(t) &= \frac{I(t)}{N} \\ R_N(t) &= \frac{R(t)}{N}, \end{aligned} \tag{7.2}$$

där $N = S + I + R$ som är den fullständiga populationen. Från (7.1) och (7.2) fås följande modell

$$\begin{cases} \frac{dS_N}{dt} = (\lambda - \mu)S_N + \lambda(I_N + R_N) - \beta S_N I_N \\ \frac{dI_N}{dt} = \beta S_N I_N - \gamma I_N - \mu I_N \\ \frac{dR_N}{dt} = \gamma I_N - \mu R_N, \end{cases} \tag{7.3}$$

där parametrarna λ , μ , β , $\frac{d\mu(t)}{dt}$ och γ är normaliserade.

Genom att numeriskt integrera (7.3) fås $S_N(t)$, $I_N(t)$ och $R_N(t)$. För att få antalet infekterade och $I(t)$ och antalet tillfrisknade $R(t)$ multipliceras $I_N(t)$ och $R_N(t)$ med N ,

$$\begin{aligned} I(t) &= \frac{I(t)}{N} \cdot N \\ R(t) &= \frac{R(t)}{N} \cdot N. \end{aligned} \tag{7.4}$$

När modellen är rätt kalibrerad kommer I och R vara precisa uppskattningar.

Andra relevanta parametrar kan även noggrant uppskattas. Man definierar

$$\begin{aligned}d(t) &= \int_{t-1}^t \frac{d\mu(t)}{dt}(\xi) \cdot I(\xi) d\xi \\D(t) &= \int_0^t \frac{d\mu(t)}{dt} \xi \cdot I(\xi) d\xi \\r(t) &= \int_{t-1}^t \gamma(\xi) \cdot I(\xi) d\xi \\i(t) &= \int_{t-1}^t \beta(\xi) \cdot S(\xi) \cdot I(\xi) d\xi\end{aligned}\tag{7.5}$$

där $d(t)$ är antalet döda fram till dag t , $D(t)$ är det ihopsamlade antalet döda fram till dag t , $r(t)$ är antalet nya fall av infektion fram till dag t och $i(t)$ är dagliga antalet fall tillfrisknade fram till dag t .

7.1 Parametrarna under första vågen av COVID-19

I det här avsnittet kommer funktionerna $\beta(t)$, $\frac{d\mu(t)}{dt}$ och $\gamma(t)$ anpassas till första vågen. Först kommer det att beskrivas metoder för att få bra uppskattningar av $\beta(t)$ och $\frac{d\mu(t)}{dt}$ för respektive område. Uppskattningen är delad i två delar. Först beräknas den dagliga infektions- och dödsgraden från data. Låt β_e och $\frac{d\mu_e}{dt}$ vara empiriska infektions- och dödlighetsgraden orsakade av COVID-19 som beräknas på följande sätt

$$\begin{aligned}\beta_e(t+1) &= \frac{i_e(t+1)}{I_e(t)} \\ \frac{d\mu_e}{dt}(t+1) &= \frac{d_e(t+1)}{I_e(t)}\end{aligned}\tag{7.6}$$

där $i_e(t)$ är antalet rapporterade nya infektioner under dag t , $d_e(t)$ är antalet rapporterade dödsfall under dag t och $I_e(t)$ är antalet rapporterade aktiva fall under dag t . Det ihopsamlade antalet dödsfall, observerade fram till dag t , kommer att betecknas med $D_e(t)$. Efter att $\beta_e(t)$ och $\frac{d\mu_e(t)}{dt}$ är bekanta kan $\beta(t)$ och $\frac{d\mu(t)}{dt}$ uppskattas med

$$\begin{aligned}\beta(t) &= \beta_0 + e^{a_\beta + b_\beta t} \\ \frac{d\mu(t)}{dt} &= \frac{d\mu_0}{dt} + e^{a_{\frac{d\mu}{dt}} + b_{\frac{d\mu}{dt}} t}\end{aligned}\tag{7.7}$$

där $b_\beta, b_{\frac{d\mu}{dt}} < 0$. Därmed antas det att parametrarna $\beta(t)$ och $\frac{d\mu(t)}{dt}$ minskar exponentiellt till deras standardvärde β_0 respektive $\frac{d\mu_0}{dt}$. Värdena a_β och b_β beräknas

med en linjär regressionsmodell med punkterna

$$(t, \log |\beta_e(t) - \beta_0|) \quad (7.8)$$

På motsvarande sätt fås $a_{\frac{d\mu}{dt}}$ och $b_{\frac{d\mu}{dt}}$ med punkterna

$$\left(t, \log \left| \frac{d\mu_e(t)}{dt} - \frac{d\mu_0}{dt} \right| \right). \quad (7.9)$$

För att bestämma γ beräknas den observerade dagliga tillfriskningsgraden $\gamma_e(t)$ som

$$\gamma_e(t+1) = \frac{r_e(t+1)}{I_e(t)}, \quad (7.10)$$

där $r_e(t)$ är antalet observerade tillfrisknade under dag t . För att anpassa modellen i respektive land kommer det att antas att γ är konstant varje kalendermånad eller under anpassningstiden av funktionerna $\beta(t)$ och $\frac{d\mu(t)}{dt}$. Värdena för γ kommer att väljas i enlighet med $\gamma_e(t)$.

7.1.1 Italien

Begränsningar för att minska spridningen av COVID-19 togs i bruk den 10 mars 2020 i Italien. Data för Italiens del börjar 10 dagar senare, den 20 mars. Denna dag betecknas som dag noll för Italiens del i studien. Först bestäms infektions- och dödsgraden under dagen t , $\beta(t)$ respektive $\frac{d\mu(t)}{dt}$, används antalet rapporterade dagliga nya fall under dag t , som betecknas med $i_e(t)$. Låt $r_e(t)$ vara rapporterade tillfrisknande under dag t och $d_e(t)$ rapporterade dödsfall under dag t . Låt även $I_e(t)$ vara antalet rapporterade aktiva fall vid tidpunkten t . Låt $R_e(t)$ och $D_e(t)$ vara totala antalet tillfrisknade respektive döda fram till dagen t . Som tidigare, ges $\beta(t)$ och $\frac{d\mu(t)}{dt}$ av (7.7). Från data kan det observeras att $\beta_0 = 0,009$ vilket minskar på avståndet mellan β_e och β . Från data kan det även observeras att $\frac{d\mu_0}{dt} = 0,0008$ vilket minskar på avståndet mellan $\frac{d\mu_e}{dt}$ och $\frac{d\mu}{dt}$. För att använda modellen (7.3) behövs λ och μ som fås från tabell 7.1. Nu behövs endast γ för att beräkna modellen för Italiens fall. Det visar sig att det månatliga medelvärdet av $\gamma_e(t)$ har stabiliserats sedan slutet av juli med värdet 0,0160. Oberoende, under de första månaderna av pandemin var värdena för $\gamma_e(t)$ generellt större. Medelvärdena för $\gamma_e(t)$ i tidsperioder, som betecknas med γ , och ses i tabell 7.2.

För att granska att modellen stämmer har man jämfört de rapporterade aktiva fallen $I_e(t)$ med resultatet från $I(t)$ under 40 dagar. Jämförelsen visar att uppskattningen är bra och kan användas [13].

Tidsperiod	Värden för γ
Kalibreringsperiod	0,0191
Maj	0,0375
Juni	0,0405
Juli	0,0233
Augusti och vidare	0,0160

Tabell 7.2: Kalibrering av γ för Italien [13].

7.1.2 Spanien

Spanska regeringen bestämde att ta i bruk begränsningar för att minska på spridningen av COVID-19 den 14 mars 2020. På samma sätt som i Italien börjar data tio dagar efter att begränsningarna togs i bruk, alltså den 24 mars. På samma sätt är det här dag noll för Spaniens del. Låt $\beta(t)$ och $\frac{d\mu(t)}{dt}$ beräknas på samma sätt som i (7.7). Passliga värden för β_0 är 0,015 och för $\frac{d\mu_0}{dt}$ är 0,00055. Låt γ vara medelvärdet av $\gamma_e(t)$ från 24 mars till 18 maj. Spanien publicerade inte data över dagliga antalet personer som tillfrisknat efter 18 maj 2020 av okänd orsak. Det här leder till att man inte kan veta hur många aktiva fall det finns i Spanien efter den 18 maj 2020.

När man vet $\beta(t)$ och $\frac{d\mu(t)}{dt}$ samt har bestämt β_0 och $\frac{d\mu_0}{dt}$ för Spanien görs en 37 dagars simulering av modellen och resultatet jämförs med data. Låt $\gamma = 0,0375$ under den 37 dagar långa undersökningsperioden. Det insamlade data passar relativt bra överens med simuleringen [13].

7.1.3 USA

Den här anpassningen börjar på samma sätt som de två tidigare. Data som används är från 2 april till 2 maj 2020. Orsaken till det är att man i de flesta delstater införde begränsningar för att minska smittspridningen av COVID-19 tio dagar före den 2 april. Låt $\beta_0 = 0,0165$ och $\frac{d\mu_0}{dt} = 0,00032$. För γ är medelvärdet för månatliga värden av $\gamma_e(t)$ homogena, som i Italien. Värdena för γ finns i tabell 7.3

För att granska att modellen stämmer har man jämfört de rapporterade aktiva fallen $I_e(t)$ med resultatet från $I(t)$. Jämförelsen visar att uppskattningen är bra och kan användas [13].

Tidsperiod	Värden för γ
April	0,0090
Maj	0,0160
Juni	0,0165
Juli	0,0185
Augusti	0,0165
September	0,0160
Oktober	0,0190
November	0,0200

Tabell 7.3: Kalibrering av γ för USA [13].

7.2 Modellens beteende under den första, andra och eventuellt tredje vågen av COVID-19

I det här avsnittet analyseras utvecklingen av COVID-19 med modellen som beskrivits tidigare. Om inga signifikanta förändringar sker gällande förhållandena stämmer $\beta(t)$, $\frac{d\mu(t)}{dt}$ och γ som definierades tidigare. Dock inträffade signifikanta förändringar i alla tre länder som undersöktes i slutet av undersökningsperioden, bland annat lättades det på begränsningarna. Det ledde till ökade antal kontakter och ökad sannolikhet att en kontakt leder till infektion.

Förändringarna påverkade inte dödligheten men de ökade spridningen av infektionen. För att simulera den andra vågen måste $\beta(t)$ kalibreras på nytt vid ett lämpligt tillfälle. Resten av parametrarna kan hållas som de tidigare definierats. Den nya funktionen för smitta kommer att betecknas med $\beta^*(t)$. Orsaken till den andra vågen var främst socialt beteende och inte miljöförändringar. Det går att se att den andra vågen började redan i juni vilket tyder på att människors beteende, som till exempel att åka på semester, påverkade att den andra vågen började.

7.2.1 Italien

I Italien kan man observera en stadig tillväxt av värdet $\beta_e(t)$ från och med juli. Värden för $\frac{d\mu_e(t)}{dt}$ har mer förutsägbart beteende än $\beta_e(t)$. Därmed, för att simulera den andra vågen av COVID-19 i Italien, kan det antas att $\frac{d\mu(t)}{dt}$ redan

har uppskattats till dess standardvärde $\frac{d\mu_0}{dt}$ till juli. Däremot måste β omformas enligt värden från och med juli $\beta_e(t)$. Låt $\beta(t)$ ersättas med $\beta^*(t)$, där $\beta(t)$ och $\beta^*(t)$ är lika enda till 16 juli ($t = 117$). För Italien beskrivs $\beta^*(t)$ av

$$\beta^*(t) = \begin{cases} \beta(t) & \text{om } t \leq 117, \\ \beta(117) + \frac{5,5\beta_0 - \beta(117)}{30}(t - 117) & \text{om } 117 < t \leq 164, \\ 5, 5\beta_0 & \text{om } 164 < t \leq 195, \\ 5, 5\beta_0 + \frac{4,5\beta_0}{15}(t - 195) & \text{om } 195 < t \leq 210, \\ 10\beta_0 & \text{om } 210 < t \leq 225, \\ 10\beta_0 - \frac{4,5\beta_0}{15}(t - 225) & \text{om } 225 < t \leq 240. \end{cases} \quad (7.11)$$

Där $t = 117$ motsvarar 16 juli, $t = 164$ motsvarar 31 augusti, $t = 195$ motsvarar 1 oktober, $t = 210$ motsvarar 16 oktober, $t = 225$ motsvarar 31 oktober och $t = 240$ motsvarar 15 november. Jämförelsen mellan $\beta(t)$ och $\beta^*(t)$ visar att uppskattningen är bra och kan användas [13].

7.2.2 USA

I USA hade redan en andra våg av COVID-19 brutit ut under sommaren, och dödligheten ökade i augusti. Den andra vågen började i juli då det sker en kraftig ökning i $\beta_e(t)$ samtidigt som smittan fördubblades. Under augusti minskade $\beta_e(t)$ och stabiliserades till sitt standardvärde β_0 i september. I början av oktober ökade $\beta_e(t)$ igen vilket skapade en tredje våg av COVID-19 i USA. Låt $\beta^*(t)$ beskriva smittan i den andra och tredje vågen,

$$\beta^*(t) = \begin{cases} \beta(t) & \text{om } t \leq 69, \\ \beta(69) + \frac{2\beta_0 - \beta(69)}{15}(t - 69) & \text{om } 69 < t \leq 84, \\ 2\beta_0 & \text{om } 84 < t \leq 120, \\ 2\beta_0 - \frac{\beta_0}{15}(t - 120) & \text{om } 120 < t \leq 135, \\ \beta_0 & \text{om } 135 < t \leq 182, \\ \beta_0 + \frac{1,5\beta_0}{40}(t - 182) & \text{om } 182 < t \leq 227. \end{cases} \quad (7.12)$$

Där $t = 69$ motsvarar 10 juni, $t = 84$ motsvarar 25 juni, $t = 120$ motsvarar 31 juli, $t = 135$ motsvarar 15 augusti, $t = 182$ motsvarar 1 oktober och $t = 227$ motsvarar 8 november. Resultatet som fås med dessa val av $\beta^*(t)$ modellerar

med bra noggrannhet det epidemiologiska läget i USA under den andra och tredje vågen [13].

7.2.3 Spanien

På grund av att Spanien inte publicerar antalet dagligen tillfrisknade personer efter 18 maj 2020 är svårt att skapa funktionen för smitta $\beta^*(t)$ under den andra vågen, eftersom det inte går att beräkna $\beta_e(t)$. Genom att använda en heuristisk process har man hittat en funktion $\beta^*(t)$ så att resultatet sammanfaller med insamlat data för kumulativa incidensen per 100 000 invånare. Då blir $\beta^*(t)$

$$\beta^*(t) = \begin{cases} \beta(t) & \text{om } t \leq 89, \\ \beta(89) + \frac{6,2\beta_0 - \beta(89)}{56}(t - 89) & \text{om } 89 < t \leq 145, \\ 6,2\beta_0 - \frac{2,2\beta_0}{16}(t - 145) & \text{om } 145 < t \leq 161, \\ 4\beta_0 & \text{om } 161 < t \leq 181, \\ 4\beta_0 - \frac{1,5\beta_0}{20}(t - 181) & \text{om } 181 < t \leq 201, \\ 2,5\beta_0 + \frac{2,1\beta_0}{15}(t - 201) & \text{om } 201 < t \leq 216, \\ 4,6\beta_0 - \frac{2,1\beta_0}{20}(t - 216) & \text{om } 216 < t \leq 236. \end{cases} \quad (7.13)$$

Där $t = 89$ motsvarar 21 juni, $t = 145$ motsvarar 16 augusti, $t = 161$ motsvarar 1 september, $t = 181$ motsvarar 21 september, $t = 201$ motsvarar 11 oktober, $t = 216$ motsvarar 26 oktober och $t = 236$ motsvarar 15 november. Fastän $\beta^*(t)$ beräknats på ett annat sätt i det här fallet, visar det sig ändå vara ett lämpligt sätt att modellera pandemin i Spanien under den andra vågen. Därför kan $\beta^*(t)$ användas för att göra uppskattningar av utvecklingen av aktiva fall $I(t)$ och antalet individer som är immuna $R(t)$.

7.3 Uppskattningar för utvecklingen av COVID-19 i Italien och USA

I det här avsnittet kommer modellen projiceras på Italiens och USA:s utveckling av COVID-19. Spaniens data är bristfälligt och därför kommer Spanien inte att beaktas i den här delen.

7.3.1 Italien

Om man analyserar data från Italien från mitten av november kan man se att nya infektioner, kumulativ incidens, daglig dödlighet och andra epidemiologiska indikatorer ökar och stabiliseras kring höga värden. Däremot har funktionen för smitta, $\beta_e(t)$, minskat sedan november i Italien. Den här minskningen kommer att påverka signifikanta epidemiologiska värden under följande veckor, börjande från 16 november 2020 ($t = 241$). För att uppskatta kommer funktionen $\frac{d\mu(t)}{dt}$ (som i det här stadiet kommer att vara samma som $\frac{d\mu_0}{dt}$) att vara samma som tidigare. Funktionen $\beta^*(t)$ kommer att utökas till $t \geq 241$,

$$\beta^*(t) = \begin{cases} 10\beta_0 - \frac{4,5\beta_0}{15}(t - 225) & \text{om } 225 \leq t < 255, \\ \beta_0 & \text{om } t \geq 255. \end{cases} \quad (7.14)$$

Observera att $t = 225$ motsvarar 15 november och $t = 255$ motsvarar 30 november.

Resultatet från modellen för daglig dödlighet och kumulativa incidensen under 14 dagar per 100 000 invånare har två olika scenarier. I det sämre scenariot antas att tillfriskningskonstanten γ har samma värde som i augusti, $\gamma = 0,0160$. I det mer optimistiska scenariot har γ samma värde som i juni, $\gamma = 0,0405$. I det sämre scenariot kommer incidensen inte under 50 före mitten av april 2021. Det här leder till att från 16 november 2020 kan ungefär 62 000 personer ha dött före mitten av april 2021. I det bättre scenariot skulle kumulativa incidensen falla under 50 i början av 2021. Då skulle 16 000 personer dö mellan 16 november 2020 och början av 2021 och ungefär 4 700 fler under resten av året 2021. Den daglig dödligheten skulle inte sjunka under 10 förrän i slutet av mars 2021, och dagliga nya fall skulle inte falla under 50 innan mitten av april 2021.

7.3.2 USA

Det kan läsas från USA:s data att den tredje vågen ökar och inte visar några tecken på att minska till mitten av november 2020. Även om en drastisk förändring skulle ske fram till mitten av november med en linjär minskning av $\beta^*(t)$ till standardvärde β_0 , på exempelvis 15 dagar, skulle dödstalet ändå vara minst 1000 per dag ända till slutet av januari 2021. För att beräkna uppskattningen

utökas $\beta^*(t)$ så att

$$\beta^*(t) = \begin{cases} \frac{43}{16}\beta_0 - \frac{9\beta_0}{80}(t - 227) & \text{om } 227 \leq t < 242, \\ \beta_0 & \text{om } t \geq 260. \end{cases} \quad (7.15)$$

Där $t = 227$ motsvarar 16 november och $t = 242$ motsvarar 30 november. Med denna smittsamhetsfunktion skulle pandemin spridas i USA i flera månader, incidens skulle vara över 100 fall ännu i mars 2021.

Kapitel 8

Avslutning

Det skulle vara enkelt att förutspå infektionssjukdomars spridning om människors beteende skulle vara begränsat. Då skulle man i vissa fall kunna använda Kermacks och McKendricks modeller för att skapa modeller av spridningen och de kunde förutspå spridningen noggrant. I de flesta länder är inte människors beteende begränsat fastän det vidtas åtgärder för att minska på spridningen av en infektionssjukdom. Infektionssjukdomen sprids ofta oberoende och det kan vara svårt att kontrollera smittspridningen.

Som det tidigare nämnts är reproduktionstalet, R_0 , ett av de viktigaste elementen inom epidemiologisk matematiska modeller. Talet berättar om förväntade antalet personer en infekterad person smittar i en mottaglig population. Även om R_0 är en viktig del i modeller är det ofta svårt att bestämma. Infektionssjukdomens beteende och hur den smittar har en betydelse för hur man ska beräkna R_0 . Även för samma infektionssjukdom finns det flera olika R_0 -värden, beroende på vilken modell man har använt [16].

Kermack och McKendricks modell publicerades 1927 och används fortfarande men med modifikationer. Den ursprungliga modellen var SIR modellen, som man har utvecklat. Modellerna som SIR , $SEIR$ och SIS har begränsade antaganden vilka sällan är realistiska i verkligheten. Det leder till att modellerna behöver anpassas när de används för att förutspå spridningen av en infektionssjukdom. Det kan man observera i kapitel 7 där parametrarna ändrades för att bättre beskriva till exempel populationsändringen.

Hur populationen mäts och definieras har en stor betydelse när man uppskattar infektionssjukdomars spridning. För att kunna bekräfta en undersökning

måste det framkomma hur populationen mäts. Det är även viktigt att parametrarna är väl beskrivna för att man ska kunna förstå och analysera modellen som använts.

Ingen person kan förutspå framtiden. Modeller som skapats för att förutspå spridningen av COVID-19 har man, som i kapitel 7, testat modellerna på existerande data för att se om modellen ger liknande resultat som data. Det förutsätter att omgivningen var infektionssjukdomen sprids och människors beteende hålls konstant i populationen även i framtiden. Det går givetvis att göra flera uppskattningar med att beakta olika omständigheter men det kan vara svårt att vidta åtgärder på basen av dem när resultatet av modellen kan vara olika.

Litteraturförteckning

- [1] Anderson, R. M (1982). *Population Dynamics of Infectious Diseases: Theory and Applications*. Springer: New York.
- [2] Auranen, K. & Leino, T. (2020). *Korona epidemian mallinnuksen lähtökohdat ja ongelmat*. Hämtad 2 juli 2021, från <https://www.duodecimlehti.fi/xmedia/duo/duo15773.pdf>.
- [3] Brauer, F. & Castillo-Chavez, C. (1994) *Basic models in epidemiology*. Hämtad 2 september 2021, från <https://ecommons.cornell.edu/bitstream/handle/1813/31835/BU-1249-M.pdf;jsessionid=4C1E66F5B0202ABF9A603A3ABC435EE8?sequence=1>
- [4] Britton, T. (2020). *Mathematics of Corona Outbreak*. Hämtad 12 januari 2021, från <https://www.youtube.com/watch?v=gSqIwXl6IjQ>.
- [5] Diekmann, O. (2019) The 1927 epidemic model of Kermack and McKendrick: a success story or a tragicomedy?. Janhunen, J. (Red.), *Sphinx 2019-2020 : vuosikirja = årsbok = yearbook* (s. 35 – 40). Societas Scientiarum Fennica 2020: Helsingfors.
- [6] Diekmann, O., Heesterbeek, H. & Britton, T. (2012). *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press: Princeton. Hämtad 13 april 2021, från https://abo.finna.fi/Record/abo_electronic_aa.9913493842905972.
- [7] Evans, A. (1982) *Viral infections of humans*. Plenum Medical Book Company: New York.
- [8] Gupta, A.K. & Nagar, D. K. (1999). *Matrix Variate Distributions*. Hämtad 28 juni 2021, från <https://www-taylorfrancis-com.ezproxy.vasa.abo>

- .fi/books/mono/10.1201/9780203749289/matrix-variate-distributions-gupta-nagar
- [9] Hethcote, H.W. (1976). Qualitative analyses of communicable disease models *Mathematical Biosciences*. Vol. 28, 335–356. Hämtad 6 september 2021, från http://people.kzoo.edu/barth/math280/articles/communicable_disease.pdf
- [10] Keshet. L. (2007). *Science 1 Problems*. Hämtad 27 september, från <https://personal.math.ubc.ca/~keshet/Sci1/chapter15Problems.pdf>
- [11] Kermack, W. O. & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society*, A115, s. 700-721. Hämtad 18 maj 2021, från <https://royalsocietypublishing.org/doi/10.1098/rspa.1927.0118>.
- [12] Lindström, M. (2020) *Linjär Algebra - Kompendium*. Åbo Akademi: Åbo.
- [13] Muñoz-Fernández, G.A., Seoane, J.M. & Seoane-Sepúlveda J.B. (2021). A SIR-type model describing the successive waves of COVID-19. *Chaos, Solitons and Fractals*. Vol. 14. Hämtad 20 oktober 2021, från <https://www.sciencedirect.com/science/article/abs/pii/S0960077921000357>
- [14] Sallet, G. (2018). *Mathematical Epidemiology*. Université de Lorraine: Lorraine. Hämtad 17 maj 2021, från <https://iecl.univ-lorraine.fr/files/2021/03/Lecture-Notes-Pretoria-2018.pdf>.
- [15] Soper, H. E. (1929). Interpretation of preiodicity in disease prevalence. *Journal of the Royal Statistical Society*, Vol. 92, 34–73. Hämtad 6 september 2021, från https://www.jstor.org/stable/2341437?seq=1#metadata_info_tab_contents
- [16] THL. (2020). *Koronavirusepidemian mallentaminen*. Hämtad 26 oktober 2021, från <https://thl.fi/documents/10531/5814325/THL+korona+mallinnuswebinaarin+kalvot+20.4.2020.pdf/cc23f5f0-87a0-8080-7b5c-121b5532bf5c?t=1587534905262>.
- [17] THL. (2021). *Koronaviruksen tarttuminen ja itämisaika*. Hämtad 25 maj 2021, från <https://thl.fi/fi/web/infektiotaudit-ja-rokotukset/a>

jankohtaista/ajankohtaista-koronaviruksesta-covid-19/tarttuminen-ja-suojautuminen-koronavirus/koronaviruksen-tarttuminen-ja-itamisaika.

- [18] THL. (2021). *Koronaepidemia: alueiden tilanne, suosituksset ja rajoitukset*. Hämtad 9 juni 2021, från <https://thl.fi/fi/web/infektiotaudit-ja-rokotukset/ajankohtaista/ajankohtaista-koronaviruksesta-covid-19/tilannekatsaus-koronaviruksesta/koronaepidemia-alueiden-tilanne-suositukset-ja-rajoitukset>.