



Matematisk Modellering av Smittspridning

Utveckling av och jämförelse mellan modeller för spridning av Covid-19

Mathematical Modelling of Disease Transmission

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers

Joakim Blomqvist

Jonas Lauri

Olle Nilsson Dahlberg

Elina Petersson

Isac Stark

Matematisk Modellering av Smittspridning

Utveckling av och jämförelse mellan modeller för spridning av Covid-19

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk fysik vid Chalmers

Joakim Blomqvist

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers

Jonas Lauri

Kandidatarbete i matematik inom civilingenjörsprogrammet Bioteknik vid Chalmers

Olle Nilsson Dahlberg Elina Petersson Isac Stark

Handledare: Philip Gerlee

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2021

Förord

Vi vill tacka vår handledare Philip Gerlee som har hjälpt oss under hela projektets gång genom att svara på frågor, vara någon att bolla idéer med och läsa igenom våra rapporter.

Vi vill även tacka våra examinatorer Ulla Dinger och Maria Roginskaya på matematiska vetenskaper som funnits tillgängliga när mer generella frågor om kandidatarbetet uppkommit.

Till sist vill vi tacka personalen på Chalmers Writing Centre som hjälpt oss med rapportstrukturen och rapportskrivningen.

Gruppens arbetsgång

Ganska tidigt efter att gruppen läst in sig om grunderna i ämnet delades arbetsgruppen in i två delar, en del som ville läsa in sig mer på utvidgningar av SIR-modellen och en del som ville utveckla en individbaserad modell. Jonas och Joakim specialiserade sig i utvidgningar av SIR-modellen och Isac, Elina och Olle specialiserade sig i individbaserade modeller. Uppdelningen ledde till att delgrupperna jobbade separat, men möten hölls regelbundet med hela gruppen för att diskutera relationen mellan modellerna och se till att arbetet blev sammanhängande.

Gruppen som fokuserat på utvidgningar av SIR-modellen började med att läsa om olika deterministiska modeller. Framst var det utvidgningar av SIR-modellen och dess kusiner som undersöktes. Ett spår som undersöktes var effekten av heterogena populationer med olika beteenden, där det generaliserades att vissa folkgrupper hade kollektivt ett beteende som medförde färre kontakter. Detta modellerades därför med olika värden på interaktionsparametern samt ett antagande att parametern var tidsberoende. Dock var detta inte särskilt intressant, eftersom effekterna i SIR-ramverket var små. Därför lades fokus därefter främst på stokastiska modeller och dess tillämpningar på smittspridning modeller.

Gruppen som fokuserat på den individbaserade modellen arbetade med att programmera nya funktioner och anpassa redan etablerade funktioner. Detta skedde genom inläsning, bolla idéer med gruppmedlemmarna och handledaren och att testa sig fram. Koden till nätverksmodellen har tagits fram gemensamt av Isac och Elina, och delvis Olle. Ansvaret har pendlat mellan oss över tiden för arbetet och vi har tillsammans stått för nya idéer och problemlösningar. Parameterestimering från litteratur delades upp så att alla fick ett par var.

Individuell arbetsgång

Dagbok och tidslogg har förts varje vecka under projektets gång. Nedan beskrivs kort gruppmedlemmarnas bidrag till arbetets resultat utöver de delar som är presenterade i tabellen Bidragsrapport för slutrapport nedan.

Jonas: Det första jag gjort efter att vi bestämde vilka modeller vi ville undersöka var fixa simuleringen av modellerna i Python. Sedan hur vi ville estimerat parametrarna till modellerna. Till detta skrev jag en algoritm för Approximate Bayesian Computation. Jag har även skrivit algoritmerna för beräkningarna av Mean Squared Error.

När det kommer till ytterligare litteraturstudier kollade jag i början på möjligheter till att implementera reglerteknik i SIR-modellerna, till exempel kan man för vaccin använda reglerteknik. Senare kollade jag även på Maximum Likelihood-metoder. Delar som dessa fick lämnas utanför rapporten på grund av omfångs begränsningar.

Joakim: Min arbetsgång har främst bestått av litteraturstudier, i början läste man många olika artiklar om Brownskrörelse på en mikroskopisk skala. För att försöka finna en modelstruktur som behandlade partiklar/individer med verktyg från statistisk fysik. Valde dock att överge det spåret och tittade istället på stokastiska processer lite mer i allmänhet. Efter Philip Gerlee länka oss till en artikel om stokastiska processer lade man mycket arbete på litteraturstudier om stokastiskaprocesser (som drevs främst av personligt intresse, eftersom det var ett nytt område inom matematiken för mig). Vilket inkluderade Brownskrörelse, Itô-analys, martingale processer, Itô-

isometri, byte av "probability measure" för att omvandla Brownska processer. Resultatet av mina litteraturstudier kan inte helt återfinnas i rapporten på grund av det begränsade omfånget. Men främst i rapporten kan det återspeglas i teori-sektionerna skrivna av mig och härledningen av Kolmogorov-ekvationen. Efter många litteraturstudier började jag sedan skriva MATLAB-kod för simulera SDE:n och Markovprocessen. Men även kod för att studera spridningen och utseendet för realiseringarna av båda processerna. Detta har resulterat i figurerna för spridning samt det slutliga figuren där samtliga modeller jämförs. Initialt kodade jag även enklare "simuleringar"/numeriska lösningar till kompartmentaliserade modeller med hetrogena och homogena populationer i Python.

Elina: Under inläsningsfasen av arbetet läste jag mycket om olika sätt att modellera individbaserade modeller. Vi sållade tillsammans ut cellulär automat och nätverk som grund. Jag gjorde egna grundläggande program för båda och vi beslutade att nätverk var mest intressant. Initialt hade jag ansvar för gradfördelning och därmed nätverksgeneratorer. Mycket tid lade jag på att bygga upp programmet, dels genom litteraturstudier och dels genom att testa sig fram själv. Att försöka använda mobilitetsdata svarade jag för. I Metoden beskrev jag `initweight`, `updateDisease`, `adjustWeight`-funktionerna. I Diskussion skrevs stycke om framtida utvecklingar och antaganden som gjorts i modellen. Då anpassning av modellen till verkligheten skulle göras ansvarade jag för ABC för nätverksmodellen. Jag tog även fram en Monte-Carlo-metod för att uppskatta parametrar men på grund av platsbrist utslöts den. Jag har gjort de schematiska grafer som finns i rapporten.

Isac: I början utav arbetet utredde jag hurvida den individbaserade modellen skulle byggas upp. Till en början kollade jag mycket på möjligheten att använda oss utav en så kallad cellulär-automata modell. Detta valdes sedermera bort då modularitet hos grafteorin ansågs vara högre, men också då vi i gruppen tyckte det lät intressant att få in den sociala aspekten i smittspridning. När vi i nätverksgruppen väl bestämt oss och en grundläggande modell tagits fram kollade jag mycket på att försöka ta ett existerande socialt nätverk från forskning. Handledaren Philip skickade en artikel som innehöll datan för ett socialt nätverk på en skola i USA, men tyvärr var dokumentationen för datan svårtolkad och inte helt applicerbar i modulen `networkx` som vi valde att använda, därav föll den delen av arbetet bort. I Teorin beskrevs genereringen av nätverk (Albert-Barabasi) samt de grundläggande strukturerna för ett matematiskt nätverk. I Metoden beskrevs `initGraph` och även basmodulen `networkx`. Under Resultatet arbetade jag med att hitta värden på parametrarna `hospitalDeathPar`, `directDeathPar` och `hospitalPar` och tog även fram grafen med vårt bästa resultat. Till följd av detta skrev jag även diskussionen om dataosäkerhet. Precis som Elina arbetade jag lite med Monte-Carlo metoden som togs bort. I slutet av arbetet strukturerade jag även upp Appendixet.

Olle: I början lärde jag mig grunderna inom programmering, då det var en viktig del för att kunna bidra till projektet. Implementering av åldersberoende var någonting som jag utforskade länge, men sen valdes detta att läggas ner, då modellen hade blivit för komplicerad, vilket ledde till att mycket arbete utfördes i onödan. I Metoden beskrevs `updateCounter`, `plotStateDistribution`-funktionerna. Parametrarna som beskriver inkubationstiden och hur länge man är smittad är jag ansvarig för att ha uppskattat från litteraturstudie och implementerat i modellen. Jag fokuserade även på de samhällliga och etiska aspekterna och den populärvetenskapliga aspekten.

Avsnitt	Delavsnitt	Skribent
Förord		Olle
Populärvetenskaplig presentation		Joakim, Jonas, Olle
Sammanfattning/ Abstract		Elina
1 Inledning		Elina, Isac
2 Teori	2.1 Deterministiska kompartmentaliserade modeller	Joakim
	2.2.1 SIR-modell med Markovkedjor i kontinuerlig tid	Jonas
	2.2.2 Dynamik uttryckt i Kolmogorov differentialekvationer	Joakim
	2.2.3 Formulering av övergångstakterna till Markovkedjemodellen	Jonas
	2.2.4 Stokastisk SIR-modell uttryckt som en Itô drift stokastisk differentialekvation	Joakim
	2.3 Individbaserad nätverksmodell	Elina, Isac
	2.4 Mean Squared Error	Jonas
	2.5 Approximate Bayesian Computation	Elina, Isac
3 Metod	3.1 Data	Elina, Olle
	3.1.1 Mobilitetsdata	Jonas
	3.2 Deterministisk SIR-modell	Joakim
	3.2.1 SIR med mobilitetsdata	Jonas
	3.3 Markovkedjemodell	Jonas
	3.4 SDE SIR-modell	Jonas, Joakim
	3.5 Nätverksmodell	Elina, Isac, Olle
4 Resultat	4.1 SIR med mobilitetsdata	Jonas
	4.2 Anpassning av stokastiska modeller	Jonas
	4.3 Spridning för SDE och Markovmodellen	Joakim
	4.4 Anpassning av nätverksmodell	Elina, Isac
	4.5 Möjligt resultat från nätverksmodell	Isac
	4.6 Kumulativa inläggningar för SDE, Markov och nätverksmodellen	Joakim, Elina, Isac
5 Diskussion	5.1 Jämförelse av de framtagna modellerna	Elina
	5.2 SIR med mobilitetsdata	Jonas
	5.3 Itô SDE:n och Markovmodellen	Joakim
	5.4 Nätverksmodellen	Elina, Isac, Olle
	5.5 Avslutning	Joakim
	5.6 Samhälleliga och etiska aspekter	Elina, Olle
Appendix	A Nästa-generationsmetoden	Joakim
	B.1 Övergångsmatrisen \mathcal{P}	Jonas
	B.2 Liten o -notation	Jonas
	B.3 Härledning av Kolmogorovs framåt differentialekvationer	Joakim
	B.4 Generatormatrisen Q	Jonas
	B.5 Gillespies algoritm	Jonas
	B.6 Härledning av τ	Jonas
	C Figur gradfördelning	Elina
	D Data	Jonas
	E Länk till koden	Elina
	F, Uppskattning av <i>hospitalPar</i> till nätverksmodellen	Elina
	G Sannolikheten att pandemin dör tidigt	Jonas
	H Spridning för Markovprocess modellen	Joakim
	I Spridning för Itô-SDE modellen	Joakim
Referenser	Alla	

Populärvetenskaplig presentation

När en sjukdom snabbt sprider sig i ett samhälle och smittar många kallas det en epidemi, men när sjukdomen börjar sprida sig över större regioner kallas det en pandemi. Pandemier har uppkommit i alla tider med exempel som digerdöden, spanska sjukan och i modernare tid Covid-19, som är anledningen till att detta arbetet utfördes. När många människor sedan blir smittade, kan det leda till att åtgärder måste tas, som att exempelvis införa restriktioner för invånarens rörelse och interaktion. För att dessa åtgärder ska vara så effektiva som möjligt krävs information om hur sjukdomen sprids, hur snabbt man blir sjuk och hur länge man är sjuk m.m. och ett viktigt verktyg som man använt sig av i flera hundraårens tid är att försöka skapa modeller som beskriver sjukdomens spridning. Med en fungerande modell kan man då planera sitt nästa steg beroende på pandemins fortsatta spridning.

Målet med arbetet var att modellera den pågående Covid-19-pandemin med olika modeller och se vilka styrkor/svagheter de olika modellerna har och om vi till slut kan skapa en modell som korrekt beskriver pandemin från dess start till idag.

Inom matematisk epidemiologi finns en rik tradition av så kallade kompartmentaliserade modeller. Dessa modeller togs fram av W. O. Kermack och A. G. McKendrick från år 1927 till 1933 och som namnet antyder så delar dessa typer av modeller in populationen i distinkta tillstånd. Denna modellstruktur är ganska mångsidig och kan bland annat tillåta studier av heterogena populationer, dock kommer homogen population antas i detta arbete. Dessa kompartmentaliserade modeller kan vara antingen deterministiska, vilket betyder att det finns ett entydigt resultat eller så kan de vara stokastiska, vilket betyder att det finns en slumpmässig faktor som skapar ett resultat med viss varians.

Vi kommer att börja med att gå igenom en enkel deterministisk modell, SIR-modellen. Denna modell bygger på att man delar upp den totala populationen i tre fack *susceptible* (S), *infected* (I) och *recovered* (R). Efter detta definieras parametrarna för populationen samt för hur sjukdomen beter sig. Detta är SIR-modellen i dess enklaste form. Man kan även utöka denna modell med fler tillstånd om man vill ta hänsyn till andra faktorer i smittspridningen. Till exempel kan inkubationsperiod, tiden mellan att du smittas och visar symptom, läggas till som ytterligare ett tillstånd. Detta är fallet för den individbaserade modellen, som är en annan modell som utvecklats.

Sedan kan stokastiska element införas till modellen eftersom det i verkligheten finns många slumpmässiga moment i smittspridningen. Vi kommer att göra detta på två olika sätt. Först med en slumpvandring i en så kallad Markovkedjmodell. Sedan med tillagd Brownsk rörelse för att få stokastiska differentialekvationer. Dessa stokastiska modeller är bra för att beskriva tidiga skeden av pandemiförlopp när sjukdomen inte är särskilt utbredd i samhället eller då små populationer studeras.

En individbaserad modell har även tagits fram där man simulerar ett samhälle med individer i ett nätverk. Namnet nätverksmodell kommer från att dessa individer är anslutna till varandra i olika grad, vilket ska representera relationerna vi har till folk runt om oss, exempelvis träffar du troligen din familj oftare än din frisör. Sedan låter man individerna träffa folk i sitt nätverk över tid och studerar hur en smittad person kommer i kontakt med sitt nätverk och hur sjukdomen sprids vidare.

För att dessa modeller ska kunna beskriva verkligheten behövs viss data för sjukdomen, exempelvis hur länge man är sjuk, dödligheten av sjukdomen m.m. Dessa faktorer finns det mycket aktuell forskning och många studier om, så för dem kunde vi få ett ganska pålitligt värde på, men vissa faktorer är svårare att sätta ett värde på, exempelvis chansen att bli smittad vid kontakt med en smittad. Genom att variera parametervärden och jämföra modellen med information om hur pandemin uppträtt sig hittills, kan man dra slutsatser om vilka värden de olika parametrarna ska ha. Detta skedde med hjälp av beräkningsmetoden "Approximate Bayesian Computation".

Efter en jämförelse av modellerna visade det sig att de hade olika styrkor/svagheter. Den deterministiska SIR-modellen anses som mest robust då den alltid ger samma resultat, till skillnad från de andra tre modellerna. Då modellerna varierar i komplexitet, tar de olika lång tid att simulera, vilket kan vara intressant om man ska köra många simuleringar. SIR-modellen var snabbast då den är enklast, följt av Markovkedjmodellen och modellen med stokastiska differentialekvationer på grund av att deras stokastiska element gör att simuleringarna tar längre tid. Nätverksmodellen tar längst tid då man studerar varje individ och har fler parametrar än de andra modellerna, vilket kräver mycket datorkraft. Att nätverksmodellen har fler parametrar gör dock att man kan gå in på mer detalj när man försöker modellera samhället, men med fler parametrar kommer fler potentiella felkällor. Så modellerna har olika styrkor och svagheter och vilken modell som är bäst beror på vilket ändamål man ska ha modellen till.

Sammanfattning

För att kunna förutspå framtida pandemier och ta väl informerade beslut i en pandemi är det viktigt att det finns matematiska modeller som beskriver smittspridningen. Syftet med projektet är att ta fram och jämföra modeller som beskriver spridning av Covid-19. I rapporten presenteras fyra olika modeller; en deterministisk SIR-modell, två stokastiska SIR-modeller och en nätverksmodell. De metoder som använts till att simulera modellerna beskrivs och parametervärden som uppskattats från litteraturdata och med hjälp av beräkningsmetoder presenteras. Modellerna jämförs sedan utifrån olika kriterier och vi beskriver deras för- och nackdelar. MSE (Mean squared error) används till att uppskatta hur bra precision de olika modellerna har vid anpassning till antal sjukhusinläggningar i Sverige och nätverksmodellen kan ge bäst anpassning.

Abstract

Mathematical models of disease transmission is important in order to predict future pandemics and make well-founded decisions during a pandemic. The purpose of the project is to develop and compare models for the transmission of Covid-19. In the report we present and compare four models; one deterministic SIR-model, two stochastic SIR-models and one network model. We present the methods used and the parameter values estimated from literature and computational methods. The models are compared and their strengths and weaknesses are presented. Mean square error is used to describe the precision of the models when predicting the levels of hospitalization in Sweden, and the network model can give the best estimation for the transmission.

Innehåll

1	Inledning	1
2	Teori	1
2.1	Deterministiska kompartmentaliserade modeller	2
2.1.1	Kermack-McKendrick SIR-modell	2
2.1.2	Beräkning av \mathcal{R}_0 för SIR-modellen	3
2.2	Stokastiska SIR-modeller	4
2.2.1	SIR-modell med Markovkedjor i kontinuerlig tid	4
2.2.2	Dynamik uttryckt i Kolmogorov differentialekvationer	4
2.2.3	Formulering av övergångstakterna till Markovkedjemodellen	5
2.2.4	Stokastisk SIR-modell uttryckt som en stokastisk differentialekvation	6
2.3	Individbaserad nätverksmodell	7
2.3.1	Nätverk	7
2.3.2	Gradfördelning	7
2.3.3	Tillämpning	8
2.4	Mean Squared Error	8
2.5	Approximate Bayesian Computation	9
3	Metod	9
3.1	Data	9
3.1.1	Mobilitetsdata	9
3.2	Deterministisk SIR-modell	10
3.2.1	SIR med mobilitetsdata	10
3.3	Markovkedjemodell	10
3.4	SDE SIR-modell	11
3.5	Nätverksmodell	12
3.5.1	Beräkningsmetod för parameterskattning	14
3.5.2	Parameterskattning från data	14
4	Resultat	14
4.1	SIR med mobilitetsdata	14
4.2	Anpassning av stokastiska modeller	15
4.2.1	Markovkedjemodellen	15
4.2.2	SDE-modellen	15
4.2.3	Sannolikhet att pandemin dör ut	15
4.3	Spridning för SDE och Markovmodellen	16
4.4	Anpassning av nätverksmodell	16
4.5	Möjligt resultat från nätverksmodell	16
4.6	Kumulativa inläggningar för SDE-, Markov- och nätverksmodellen	17
5	Diskussion	17
5.1	Jämförelse av de framtagna modellerna	18
5.2	SIR med mobilitetsdata	18
5.3	Itô SDE:n och Markovmodellen	19
5.4	Nätverksmodellen	19
5.5	Samhälleliga och etiska aspekter	20
5.6	Avslutning	20
A	Nästa-generationsmetoden	24

B	Markovkedjor	25
B.1	Övergångsmatrisen \mathcal{P}	25
B.2	Liten o -notation	25
B.3	Härledning av Kolmogorovs framåt differentialekvation	25
B.4	Generatormatrisen Q	26
B.5	Gillespies algoritm	27
B.6	Härledning av τ	27
C	Figur gradfördelning	28
D	Data	29
E	Länk till koden	30
F	Uppskattning av <i>hospitalPar</i>	31
G	Sannolikheten att pandemin dör tidigt	32
H	Spridningen för Markovprocess modellen plottat för några realiseringar	33
I	Spridningen för Itô-SDE modellen plottat för några realiseringar	34

1 Inledning

Den moderna epidemiologin har sin början i 1800-talets England där läkaren John Snow kartlade spridningen av kolera i London genom att studera antalet smittade i förhållande till avståndet till vattenbrunnar och på så sätt kunde begränsa smittans fortskridande genom att isolera de påverkande brunnarna [1]. Lite mer än hundra år senare, år 1927, publicerade Kermack och McKendrick en artikel där den så kallade SIR-modellen presenterades. Modellen har sedan sin publicering utvecklats ytterligare och haft stor inverkan på hur smittspridning modelleras [2].

SIR-modellen kommer beskrivas mer detaljerat i Teori-avsnittet, men kortfattat delas befolkningen i tre tillstånd: S *suseptible*, I *infected* och R *recovered* och flöden mellan dessa tillstånd beskrivs av differentialekvationer. Modellen är en av de mest använda modellerna till att beskriva smittspridning av olika infektionssjukdomar och har än så länge en mycket central roll i dagens epidemiologi. Bland annat används den i dagsläget för att beskriva spridning av Covid-19 under den nuvarande pandemin. För att beskriva smittspridningen i Sverige använder Folkhälsomyndigheten en variant av Kermack och McKendricks väl beprövade modell som kallas SEIR-modellen. Utöver de tidigare nämnda tillstånden används tillståndet E *exposed*, vilket beskriver de individer som inte än utvecklat symptom [3].

I och med den kraftiga teknikutveckling världen sett de senaste hundra åren finns det nu nya möjligheter att förutspå och förebygga smittspridning i samhället. Bland annat har den ökade datorkapaciteten lett till att det är möjligt att utföra mer omfattande beräkningar och simulera smittspridning på andra sätt. Så kallade individbaserade modeller är en grupp modeller som kräver större datorkapacitet, men som samtidigt erbjuder stora möjligheter till att följa de verkliga mekanismerna som ligger till grund för spridning. Till exempel använde Storbritannien sig bland annat av en individbaserad modell då beslut togs om olika kontrollåtgärder för att minska smittspridning under början av Covid-19-pandemin [4].

Syftet med projektet är utveckla och jämföra matematiska modeller för smittspridning och i denna rapport presenteras och jämförs de modeller vi tagit fram. Vi kommer att redogöra för ytterligare utvidgningar av SIR-modellen och även studera effekten av att införa stokastiska element i modellen. Till skillnad från en deterministisk SIR-modell kan en stokastisk ge olika resultat med samma variabler. Detta kan vara gynnsamt då man vill studera om det går att stoppa smittspridningen från att bli en pandemi. Den stokastiska utvidgningen kommer göras på två sätt, för det första kommer en tidskontinuerlig Markovkedja användas och för det andra kommer Itô-drift tillämpas i stokastiska differentialekvationer. Förutom dessa modeller kommer vi även presentera en individbaserad nätverksmodell och slutligen jämföra dessa modeller med varandra. Nätverk är ett sätt att tillämpa individbaserad modellering på som vi valt att fokusera på för att det har egenskapen att man kan härma faktiska sociala nätverk, vilka är grunden till smittspridning.

Möjligheten att kunna förutspå antalet smittade i samhället skulle vara till stor nytta för vårdgivare för att förutspå till vilken grad sjukhusystemet kommer bli belastat, men det kan även vara av intresse för beslutsfattare för att avgöra i vilken takt nedstängningar och framtida öppningar av samhället ska ske. Förbättrade möjligheter att modellera smittspridning skulle även kunna underlätta vaccinationsprogram genom att till exempel detektera vilka grupper som löper störst risk att bli smittade. Dessutom kan modeller av smittspridning ge viktig indikation till mekanismerna bakom och beskriva vad som driver spridningen.

2 Teori

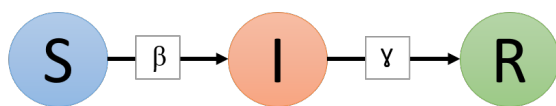
Det finns många olika modeller för att beskriva smittspridning och mängden nya publikationer ökar i takt med den nuvarande pandemin. Vi kommer här beskriva den teori som ligger bakom de modeller vi valt att fokusera på. Vi börjar med den deterministiska SIR-modellen och fortsätter sedan till att studera två sätt för hur stokastiska element kan inkorporeras. Dessa modeller antar att befolkningen är homogen och alla individer möter alla lika mycket, därför kommer vi avslutningsvis presentera hur man kan använda nätverk och grafteori till att modellera smittspridning.

2.1 Deterministiska kompartmentaliserade modeller

En välbeprövad metod för att beskriva hur sjukdomar sprider sig i en population är de så kallade kompartmentaliserade modellerna. För dessa modeller delas populationen in i olika tillstånd utifrån vilken sjukdom som beskrivs. Nedan kommer vi ta upp några viktiga deterministiska modeller och deras egenskaper.

2.1.1 Kermack-McKendrick SIR-modell

Kermack och McKendricks modell som vi nämnde i inledningen har tre olika klasser; de som kan insjukna S *susceptible*, de som är sjuka I *infected* och de som tidigare insjuknat R *removed/recovered*, vilka sedan antas ha en immunitet mot sjukdomen. Dessa klasser utgör därmed diskreta tillstånd för varje individ i populationen. Figur 1 visar hur flödet mellan de olika grupperna sker och bestäms bland annat av parametrarna β och γ .



Figur 1: En illustration av SIR-modellens tre tillstånd samt flödet mellan dem. Första parametern β är en interaktionsparameter för smittspridningen och den sista γ beskriver tillfriskningstakten.

Ett viktigt grundantagande för dessa modeller är att den totala populationen N , antas vara konstant, $N = S(t) + I(t) + R(t)$ alltså är $dN/dt = 0$. Det är ett ganska rimligt antagande för en stor population under ett kort tidsintervall, om sjukdomen har låg fatalitet och att andra faktorer som ordinarie tillökningar och bortfall från populationen kan försummas.

För att kunna modellera hur en individ rör sig mellan dessa tillstånd krävs en modell för interaktionen mellan individerna i samhället. För SIR-modeller och dess utvidgningar antas populationen beskrivas av en analog situation från kinetisk gasteori. För det två-dimensionella fallet beskrivs gaspartiklarna som slumpvandrande cirklar i ett plan. Inom SIR-modellen antar slumpvandrarerna istället rollen som enskilda individer vilka krockar med varandra, kollisionerna är mekanismen som ger upphov till smittspridningen.

Vidare antas individerna i samhället vara homogent fördelade och interaktionerna mellan dem är ej reglerade, därmed är sannolikheten att en mottaglig individ möter en infekterad individ $\mathcal{P}(t) = I(t)/N$. Varje interaktion är oberoende, vilket medför att individen rör sig enligt en opåverkad Brownsk rörelse (mer detaljer följer i Teori-sektionen om stokastiska modeller) och träffar därmed en infekterad person med sannolikheten \mathcal{P} . Varje individ eller gaspartikel har samma medel-kollisionstakt k ; $[k] = [\text{tidsenhet}^{-1}]$. Vid varje interaktion sprids sjukdomen vidare med en sannolikhet p . Vi definierar sedan en interaktionsparameter β , som: $\beta := pk$. Genom att kombinera sannolikheten att möta en infekterad individ med medeltalet för antalet interaktioner multiplicerat med sannolikheten att sprida sjukdomen för samtliga mottagliga personer $S(t)$, kan ett uttryck för övergångstakten från S till I bildas:

$$\frac{dS}{dt} = -\beta \underbrace{\frac{I(t)}{N}}_{:=\lambda} S(t) = -\lambda(t)S(t). \quad (1)$$

Vidare, så länge sjukdomen som studeras inte är kronisk, borde även individer tillfriskna i en känd takt γ , $[\gamma] = [\text{tidsenhet}^{-1}]$. Med detta i åtanke kan förändringshastigheten av antalet individer i tillståndet I beskrivas av differensen mellan de som insjuknar och de som tillfrisknar:

$$\frac{dI}{dt} = \lambda(t)S(t) - \gamma I(t). \quad (2)$$

Slutligen utgör tillfriskningstakten utökningen av individer till R tillståndet,

$$\frac{dR}{dt} = \gamma I(t). \quad (3)$$

Modellernas dynamik beskrivs därmed av ett system ickelinjära kopplade ordinära differentialekvationer (ODE), vilka kan sammanfattas som:

$$\begin{cases} \frac{dS}{dt} = -\lambda(t)S(t) \\ \frac{dI}{dt} = \lambda(t)S(t) - \gamma I(t) \\ \frac{dR}{dt} = \gamma I(t) \end{cases} \quad (4)$$

Denna typ av modellstruktur medför att det enkelt kan införas fler tillstånd på bekostnad av fler differentialekvationer. Det enda som krävs är ett uttryck för övergångstakten mellan tillstånden. Något som förenklar situationen för SIR-modellen är att endast S - och I -tillstånd krävs för att fullständigt beskriva systemet. Detta följer från antagandet om konstant population, $N = S + I + R \Leftrightarrow R = N - S - I$. Således är den sista differentialekvationen som beskriver dR/dt onödig ur ett simulationsperspektiv.

Genom en noggrannare analys av SIR-modellens differentialekvationer kan det basala reproduktionstalet \mathcal{R}_0 definieras. Det finns flera vägar att härleda detta tal. Bland annat kan det göras med en algebraisk behandling av ekvationerna och att studera kravet för att tillstånden i modellen skall vara avtagande eller växande. För mer komplicerade modeller blir därför en sådan behandling mer komplicerad.

Flera andra tillvägagångssätt har utforskats i tidigare litteratur [2][5]. Inkörsporten för intuitionen för reproduktionstalet kommer från att tänka på smittspridningen som en förgreningsprocess. Talet ska beskriva medeltalet av sekundära infektioner som uppstår från en infekterad individ. Det kan uttryckas som en integral

$$\mathcal{R}_0 = \int_0^\infty b(\tau)F(\tau)d\tau, \quad (5)$$

där, $b(\cdot)$, definieras som antalet nya infekterade som en individ producerar i medel under tiden τ . Den sista termen, $F(\cdot)$, är en funktion för sannolikheten att en nyligen infekterad individ förblir smittsam under tiden τ [5].

2.1.2 Beräkning av \mathcal{R}_0 för SIR-modellen

En annan metod vilken kommer att brukas för detta arbetet är nästa-generationsmetoden. Denna metod bygger på konstruktionen av matriser som definierar en matrisoperator känd som nästa-generationsmatrisen [5]. Se Appendix A för en beskrivning. Tillståndet med infekterade individer är endast I , därmed är $x = I$ en skalär.

$$\begin{aligned} \mathcal{F} &= \beta \frac{I(t)}{N} S(t), & \mathcal{V} &= \gamma I(t), \text{ se figur 1.} \\ F &= \frac{\partial \mathcal{F}}{\partial x} = \frac{\beta}{N} S_s, & V &= \frac{\partial \mathcal{V}}{\partial x} = \gamma. \end{aligned} \quad (6)$$

Steady-state vid det sjukdomsfria tillståndet för de motagliga individerna är $S_s = N$, därmed blir operatorn:

$$G = FV^{-1} = \frac{\beta}{\gamma}. \quad (7)$$

Avslutningsvis, det dominerande egenvärdet för G utgör värdet för \mathcal{R}_0 . Då G är en skalär innebär detta att funktionsbasen $\{1\}$ utgör en egenfunktion för operatorn med egenvärdet \mathcal{R}_0

$$\begin{aligned} A\mathbf{v} &= \lambda\mathbf{v} \\ G \cdot 1 &= \frac{\beta}{\gamma} \cdot 1 \Rightarrow \mathcal{R}_0 = \frac{\beta}{\gamma}. \quad \square \end{aligned} \quad (8)$$

Det basala reproduktionstalet kan användas som ett mått för sjukdomens spridningspotential. För $\mathcal{R}_0 > 1$ har sjukdomen potential att snabbt sprida sig i populationen. Medan $\mathcal{R}_0 < 1$ leder till att individer tillfrisknar snabbt, vilket resulterar i att möjligheten för dessa individer att infektera andra är låg, så länge $\gamma > \beta$. Tröskelvärdet $\mathcal{R}_0 = 1$ ligger därmed i gränsen för att sjukdomen skall kunna spridas effektivt i samhället [2].

2.2 Stokastiska SIR-modeller

Förutom deterministiska modeller kan även stokastiska versioner av SIR-modeller implementeras. Syftet med att införa dessa är framför allt att undersöka hur många realiseringar som har ett pandemiutfall. Eftersom smittspridningen dels leds av stokastiska processer kan antalet infekterade gå mot noll i ett tidigt stadi. Generellt är stokastiska modeller lämpliga att använda när antalet smittade är lågt, eller då populationen är liten. För dessa situationer kommer effekten av slumpen att vara mer framträdande. Modellerna kommer baseras på Markovkedjor i kontinuerlig tid och så kallade Itô stokastiska differentialekvationer. Med hjälp av sannolikheterna för smitta och återhämtning kan en stokastisk modell skapas.

2.2.1 SIR-modell med Markovkedjor i kontinuerlig tid

Den första stokastiska modellen som vi kommer att beskriva är en modell baserad på Markovkedjor i kontinuerlig tid tillsammans med ett diskret tillståndsrum för $S(t)$, $I(t)$ och $R(t)$ från SIR-modellen. Markovkedjor är ett sätt att simulera slumpvandring, för en kort sammanfattning, se Appendix B.1. Först en definition av Markovkedjor i kontinuerlig tid.

Definition 1 *En stokastisk process, bestående av en sekvens tidsberoende slumpvariabler $\{X(t)\}_{t \geq 0}$, med diskreta tillståndsrum S , är en Markovkedja i kontinuerlig tid om*

$$\mathbb{P}(X_{t+d} = x_{t+d} | X_d = x_d, X_u = x_u) = \mathbb{P}(X_{t+d} = x_{t+d} | X_d = x_d) \quad (9)$$

är uppfyllt, där $t, d \geq 0$ och $0 \leq u < d$ [6].

Detta innebär att två tillstånd x_d, x_u vid tiderna $u < d$ ej kan förutsäga framtida dynamik bättre än endast tillståndet x_d . Processer som uppfyller definition 1, innehar Markovegenskapen, vilket är att processen glömmet vilken väg den tog till det nuvarande tillståndet. Dessa tillståndsförflyttningar är således oberoende stokastiska processer. För Markovkedjorna spelar alltså vägen till det nuvarande tillståndet inte någon roll, utan endast det tidigare tillståndsvärdet. Tillståndsrummen för $S(t)$, $I(t)$ och $R(t)$ definieras enligt

$$S(t), I(t), R(t) \in \{0, 1, \dots, N\}, \quad t \in [0, t_{max}] \quad (10)$$

där $t_{max} \in \mathbb{R}$ är någon godtycklig maximal tidpunkt för simuleringen. Låt nu s och i beteckna de nuvarande värdena för $S(t)$ och $I(t)$. Då blir övergångssannolikheterna för en liten tidsförskjutning, $\Delta t > 0$,

$$\mathcal{P}_{(s,i) \rightarrow (s+k, i+j)}(\Delta t) = \begin{cases} \beta i \frac{s}{N} \Delta t + o(\Delta t), & (k, j) = (-1, +1) \\ \gamma i \Delta t + o(\Delta t), & (k, j) = (0, -1) \\ 1 - (\beta i \frac{s}{N} + \gamma i) \Delta t + o(\Delta t), & (k, j) = (0, 0) \\ o(\Delta t), & \text{annars} \end{cases} \quad (11)$$

vilka kan jämföras med SIR-modellen [7]. Istället för differentier har vi Δt -termer. För en sammanfattning av vad $o(f(x))$ notationen innebär, se Appendix B.2. Det viktiga att lägga märke till här är fallen $(k, j) = (-1, +1)$ samt $(k, j) = (0, -1)$ vilka motsvarar sannolikheten för smitta respektive återhämtning. Dessa medför att $(s, i) = (n, 0)$, $n \in \mathbb{N}$, är ett absorberande tillstånd. Ett absorberande tillstånd är ett tillstånd från vilket ytterligare förändring är omöjlig. I vårt fall är detta noll smittade. Det krävs smittade för att sjukdomen skall kunna fortsätta spridas.

2.2.2 Dynamik uttryckt i Kolmogorov differentialekvationer

Antag först att begynnelsestillståndet är $a = (s, i)$ och det nya tillståndet är $b = (s+k, i+j)$. Sedan, givet uttryck för övergångssannolikheter kan en differentialekvation för sannolikhetsfördelningarna för varje övergång $\mathcal{P}_{a \rightarrow b} = \mathcal{P}_b$, uttryckas generellt med en framåt-formulering av Kolmogorovs differentialekvationer [7]

$$\frac{d\mathcal{P}_b}{dt} = \sum_{k \neq a} (\mathcal{P}_k(t) q_{k \rightarrow b}) - q_{a \rightarrow a} \mathcal{P}_a(t). \quad (12)$$

Där $q_{i \rightarrow j}$ termerna är övergångstakterna, för detta fall står de uttryckta i ekvation (11). Varje term i summan korresponderar mot en form av övergång $S \mapsto I$, $I \mapsto R$ under ett tidsintervall Δt . Den sista termen utanför summan motsvarar sannolikheten för tillståndet att förbli oförändrat. I Appendix B.3, är en härledning av den specifika Kolmogrovekvationen för detta fall givet övergångssannolikheterna i ekvation (11) presenterad.

Resultatet kan sammanfattas som,

$$\frac{d\mathcal{P}_{(s,i)}}{dt} = \mathcal{P}_{(s+1,i-1)}(t) \frac{\beta}{N} (s+1)(i-1) + \mathcal{P}_{(s,i+1)}(t) \gamma (i+1) - \mathcal{P}_{(s,i)}(t) \left[\frac{\beta}{N} si + \gamma i \right]. \quad (13)$$

Denna ekvationen beskriver tidsutvecklingen av förändringen av sannolikheterna för tillståndsövergångarna $S \rightarrow I$ och $I \rightarrow R$. Att formulera analytiska lösningar för en sådan ekvation är inte trivialt och kräver därmed i detta arbete en numerisk hantering för att simulera pandemiförloppet.

Kolmogrovekvationen kan även uttryckas med matriskalymetoder. Nedan följer formuleringen för Kolmogrov framåt differentialekvationen

$$\frac{d\mathcal{P}}{dt}(t) = \mathcal{P}(t)Q \quad (14)$$

Matrisen Q är den så kallade infinitesimala generatormatrisen och \mathcal{P} är en matris vars element utgörs av övergångssannolikheter. Dimensionerna för matriserna Q och \mathcal{P} beror på antalet möjliga kombinationer av tillstånd som kan uppnås [7]. Några exempel på tillstånd som kan uppnås är $(s, i) \in \{(N, 0), (N-1, 1), \dots, \}$. SIR-modellstrukturen påtvingar några krav på vilka tillstånd som kan existera. Mer specifikt antagandet om en konstant population. Detta innebär exempelvis att tillstånd som (N, N) inte kan uppnås då det totala antalet individer som krävs överstiger populationens storlek.

2.2.3 Formulering av övergångstakterna till Markovkedjemoellen

För att simulera krävs takten för förflyttning från ett tillstånd till ett annat. Därmed krävs en definition av övergångstakten definierad av övergångssannolikheterna.

Definition 2 *Antag att $\mathcal{P}(t)$ är deriverbar, då är $Q := \mathcal{P}'(0)$, där Q är den infinitesimala generatormatrisen. Elementen i denna matris Q är då övergångstakterna [6].*

För mer information om Q se Appendix B.4. Börjar med att räkna ut $\mathcal{P}'(\Delta t)$

$$\mathcal{P}'(\Delta t) = \begin{cases} \beta i \frac{s}{N} + \frac{d}{dt}(o(\Delta t)), & (k, j) = (-1, +1) \\ \gamma i + \frac{d}{dt}(o(\Delta t)), & (k, j) = (0, -1) \\ 0 + \frac{d}{dt}(o(\Delta t)), & (k, j) = (0, 0) \\ \frac{d}{dt}(o(\Delta t)), & \text{annars.} \end{cases} \quad (15)$$

Vad är då $\frac{d}{dt}(o(\Delta t))$ när $\Delta t \rightarrow 0^+$? Om $f(\Delta t) = o(f(\Delta t))$ då är $\lim_{\Delta t \rightarrow 0^+} \frac{f(\Delta t)}{\Delta t} = 0$ per definition, se Appendix B.2. Dessutom gäller följande relationer

$$\frac{d}{dt} \left(\lim_{\Delta t \rightarrow 0^+} \frac{f(\Delta t)}{\Delta t} \right) = \lim_{\Delta t \rightarrow 0^+} \frac{\frac{df(\Delta t)}{d\Delta t}}{1\Delta t^{1-1}} = \lim_{\Delta t \rightarrow 0^+} \frac{df(\Delta t)}{d\Delta t} = 0 \quad (16)$$

där den första likheten kommer från L'Hôpitals regel och sista likheten kommer från definitionen ovan. Vi förutsätter då att f uppfyller förutsättningarna för att kunna använda L'Hôpitals regel. Ifall derivatan existerar är den alltså lika med noll. Detta medför att

$$\mathcal{P}'(0) = Q_{(s,i) \rightarrow (s+k,i+j)}(\Delta t) = \begin{cases} \beta i \frac{s}{N}, & (k, j) = (-1, +1) \\ \gamma i, & (k, j) = (0, -1) \\ 0, & (k, j) = (0, 0) \\ 0, & \text{annars.} \end{cases} \quad (17)$$

I övergångstakterna så försvinner beroendet på Δt .

Nu tillbaka till syftet med att införa denna modell, vilket är att undersöka om det är möjligt att stoppa sjukdomen från att bli till en pandemi. För denna analys finns det teori att tillgå men vi kommer helt enkelt bara att simulera för olika parametervärden. Speciellt för olika antal smittade vid tid noll, I_0 .

2.2.4 Stokastisk SIR-modell uttryckt som en stokastisk differentialekvation

Ett annat sätt att implementera en stokastisk SIR-modell är att använda stokastiska differentialekvationer vilka baseras på Itôprocesser. Istället för diskreta tillståndsrum och kontinuerlig tid kommer dessa använda både kontinuerliga tillståndsrum och tidssteg. Dock krävs en tidsdiskretisering för simulationen, av samma skäl som för Runge-Kuttametoden.

Vi börjar med en beskrivning av Itôprocesser. Stokastiska differentialekvationer (SDE) baserade på Itôprocesser \mathbf{X} kan generellt uttryckas med en driftterm $\boldsymbol{\mu}(\mathbf{X}, t)$ samt en variansmatris G_t och en Wienerprocess term \mathbf{W}_t :

$$d\mathbf{X} = \boldsymbol{\mu}(\mathbf{X}, t)dt + G_t d\mathbf{W}_t. \quad (18)$$

Weinerprocesser har flera bra egenskaper som exempelvis kvadratisk varians. För en SIR-modell blir slumpvariabelvektorn $\mathbf{X}(t) = [S(t), I(t)]^\top$. För att anlända i en stokastisk differentialekvation utgår vi från Markovprocess-modellen. Generellt kan en Markovprocess approximativt beskrivas av en Fokker-Planck-ekvation [8]. Fokker-Planck-ekvationen är en kontinuitetsekvation,

$$\frac{d\rho}{dt} - \nabla \cdot \mathbf{J} = 0 \quad (\rho \sim \text{rumstäthet}, \mathbf{J} \sim \text{strömtäthet}), \quad (19)$$

vilket beskriver rums- och tidsvariationer för flöden av en bevarad storhet. För stokastiska processer innebär detta att totala sannolikheten är bevarad.

En Fokker-Planck-ekvation som beskriver tidsutvecklingen av sannolikhetsfördelningen \mathcal{P} för n -tillståndsvariabler, $\mathbf{x} = [x_1, x_2, \dots, x_n]^\top$, kan uttryckas som [9]

$$\frac{\partial \mathcal{P}}{\partial t}(\mathbf{x}) = - \sum_{k=1}^n \frac{\partial}{\partial x_k} (\mathcal{P}(\mathbf{x}) a_k(\mathbf{x})) + \frac{1}{2} \sum_{j=1}^n \sum_{k=1}^n \frac{\partial^2}{\partial x_k \partial x_j} (\mathcal{P}(\mathbf{x}) b_{jk}(\mathbf{x})). \quad (20)$$

Termerna x_j definieras av värden för antalet individer i ett tillstånd x_i för en viss tidpunkt normaliserad mot totala populationen N , $x_j = x_i(t)/N$. Termerna i Fokker-planck-ekvationen svarar mot drift och varianstermerna i Itô-SDE:n:

$$\begin{aligned} \boldsymbol{\mu} = a_k(\mathbf{x}) &= \sum_{j=1}^n (T_{jk}(\mathbf{x}) - T_{kj}(\mathbf{x})), \\ G_t = b_{jk}(\mathbf{x}) &= \frac{1}{N} \left[-T_{jk}(\mathbf{x}) - T_{kj}(\mathbf{x}) + \delta_{jk} \sum_{l=1}^n (T_{jl}(\mathbf{x}) - T_{lj}) \right]. \end{aligned} \quad (21)$$

I ekvationerna för drift- och varianstermerna så motsvarar T_{ij} sannolikheterna för att en tillståndsändring sker och δ_{jk} är Kroneckerdeltatensorn. För stora populationer kommer stegen som Markovprocessen tar bli små och således medföra att Fokker-Planck-ekvationen beskriver populationsutvecklingen bättre. Sedan genom att använda verktyg från Itô-analysen kan en så kallad Langevinekvation eller Itô-drift ekvation härledas från Fokker-planck-ekvationen, då det finns en korrespondens mellan stegen i Markovprocessen och Wienerprocessen W [9],

$$d\mathbf{X} = \mathbf{a}(\mathbf{x})dt + b(\mathbf{x})d\mathbf{W}(t). \quad (22)$$

Brustermerna motsvarar Wienerprocesser, vilka för små tidssteg är $(W_{t+1}^{(j)} - W_t^{(j)}) \sim \mathcal{N}(0, \sqrt{\Delta t})$.

Mer konkret för detta arbete utgörs det kontinuerliga tillståndsrummet av $S(t), I(t)$ och $R(t)$.

$$S(t), I(t), R(t) \in [0, N].$$

Dock krävs även här endast två tillstånd för att beskriva systemet eftersom $R = N - S - I$. Itô-drift-ekvationerna som kommer användas för detta arbete är

$$dS(t) = -\frac{\beta S(t)I(t)}{N}dt - \sqrt{\frac{\beta S(t)I(t)}{N}}dW_1(t) \quad (23)$$

$$dI(t) = \left(\frac{\beta S(t)I(t)}{N} - \gamma I(t)\right)dt + \sqrt{\frac{\beta S(t)I(t)}{N}}dW_1(t) - \sqrt{\gamma I(t)}dW_2(t) \quad (24)$$

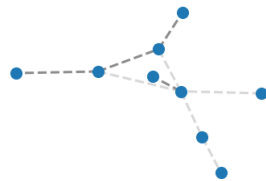
där dW_1 och dW_2 beskriver den infinitesimala differensen för två oberoende Wienerprocesser [7].

2.3 Individbaserad nätverksmodell

Individbaserade modeller, även kallade agentbaserade modeller, tar hänsyn till interaktion mellan individer genom att studera interagerande agenter i en miljö. Det finns många olika sätt att göra detta på och inom litteraturen finns ingen entydig definition på vad en agent är, men generellt gäller att agenterna har attribut, är självständiga och deras beteende styrs av bestämda regler [10]. En individbaserad modell kan exempelvis göras med hjälp av s.k. cellulär automat, där agenterna går runt på ett schackbräde och då de hamnar på samma ställe sker en interaktion. Ett annat exempel är att modellera ett nätverk av agenter, där agenterna är stationära och interaktioner sker genom sammanlänknings mellan dem [11].

2.3.1 Nätverk

Internet, neurala nätverk, näringskedjor, nätverk av affärsrelationer och sociala nätverk är några exempel på vad man kan använda nätverksmodeller till att beskriva. Ett nätverk kan visualiseras med hjälp av grafteori. En graf G består av noder i, j och kanter G_{ij} som binder samman noderna, se figur 2.



Figur 2: En graf med nio noder och kanter som sammanbinder dem. Kanternas olika nyanser symboliserar att kanterna har olika vikter.

För att modellera smittspridning används ofta sociala nätverk där en individ, eller agent, representeras av en nod och interaktioner mellan individerna är kanterna i grafen [12]. Vid mer heterogena kontaktmönster kan man alltså utnyttja nätverksmodeller för att beskriva populationen på individnivå. En faktor som kan vara av intresse vid beskrivning av ett socialt nätverk är att olika individer ofta interagerar olika mycket med olika individer i sitt kontaktnät. Detta kan inom grafteori beskrivas genom att ge varje kant mellan noderna en vikt [13]. Kanterna till grafen G_{ij} kan beskrivas av angränsningsmatrisen, där varje kant är

$$G_{ij} = G_{ji} = \begin{cases} v & \text{om det finns en kant mellan } i \text{ och } j \\ 0 & \text{annars} \end{cases} \quad (25)$$

där $v > 0$ är ett tal som anger vikten av kanten, G noterar nätverket och i och j är nod-index, således noterar G_{ij} kanten mellan nod i och nod j . I grafteori kan man definiera en självkopplad kant G_{ii} , men i smittspridningssyfte kan koppling $G_{ii} = 0$ anses vara rimligt då det ej finns någon fysikalisk innebörd i att smitta sig själv. Graden av en nod k_i är antal kanter som är kopplade till noden [13].

2.3.2 Gradfördelning

Kontaktmönstret i ett samhälle kan beskrivas med fördelning av grader, det engelska begreppet *Degree distribution* förekommer ofta inom litteraturen och innebär hur antal grader är fördelade över noderna. Vi kommer ta upp två sätt för hur gradfördelningen kan erhållas. Den första metoden är att man slumpmässigt ger varje nod en grad utifrån en fördelning, t.ex. Poissonfördelning

eller potenslagsfördelning. Det andra sättet man kan skapa en realistisk fördelning av grader på är att använda demografisk data [14], [15]. Detta tillvägagångsätt kräver att data som används är representativ för situationen som ska modelleras, t.ex. om ett socialt nätverk över Göteborg ska inrättas måste data för just Göteborg användas, vilket är en begränsning.

Låt oss fokusera mer på grafer som skapas utifrån olika fördelningar. Att ha en graf där fördelningen av grader följer en binomialfördelning är den mest studerade och har bland annat namnet Erdős-Renyi-graf. Man låter kanter skapas mellan n noder med en sannolikhet p , vilken ges av kvoten mellan genomsnittliga graden och antal noder. När storleken på populationen blir stor kommer fördelningen närma sig en Poissonfördelning, se figur 10 i Appendix C. Detta kan fungera bra för sociala nätverk där interaktionen mellan individer är homogen.

En Erdős-Renyi-graf är slumpmässig, men ofta förekommer det strukturer i verkliga sociala nätverk som man kan vilja beskriva. Det är vanligt för verkliga sociala nätverk att beskrivas av en potenslagsfördelning, vilken har en mer utdragen svans [13]. De noder som har hög grad, vilka är långt till höger i potenslagsfördelningen, kan ses som superspridare [16]. Många artiklar har även kommit fram till spridningen av Covid-19 följer en potenslagsfördelning [17], [18], [19].

En modell för att generera sådana grafer är framtagen av Albert Barabási. Nätverket börjar med en storlek på m_0 noder, varefter en ny nod genereras som bildar en kant till $m \leq m_0$ noder där m_0 och m är godtyckliga heltal, sannolikheten för en ny nod att bilda en kant med en existerande nod är proportionell till den existerande nodens grad [20]. Detta förhållande kallas preferentiell anknnytning (från engelskans *preferential attachment*). Sannolikheten att en ny nod har en kant till en befintlig nod i ges av

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \quad (26)$$

Denna anknynings-preferens bidrar till att väldigt många noder har en grad som är runt m_0 och ett fåtal noder kommer ha en mycket hög grad, då ekvationen ovan bidrar till att de som redan har många kanter får ännu fler [21]. I figur 10 i Appendix C visas fördelning för en Barabási-Albert-graf där fem noder läggs till varje ny nod vid genereringen av nätverket.

För att beskriva mer komplexa sociala nätverk kan en kombination av gradfördelning och viktning användas [22]. En nods grad kan ses som vilka individen har möjlighet att interagera med och vikten av en kant representerar hur ofta individen träffar dem. Detta kan göra att det är möjligt att simulera en ändring av socialt beteende och beskriva ett socialt nätverk mer detaljerat.

2.3.3 Tillämpning

Till skillnad från andra individbaserade modeller som t.ex. slumpvandringens modellen cellulär automat kan noderna i ett nätverk vara fixerade och därmed inte förflytta sig mellan olika positioner. Spridning av en sjukdom sker då genom att två noder har en kant mellan sig och en slumpvariabel avgör om smitta sker eller inte. Noderna delas in i olika tillstånd enligt SIR-modellen och smitta överförs mellan noderna med en viss sannolikhet. Detta kan göras på liknande sätt för både en Erdős-Renyi och Barabási-Albert-graf.

Det är även möjligt att ha ett nätverk där noder istället ses som platser istället för individer. Noderna kan exempelvis representera skolor eller affärer och individer vandrar mellan dessa och blir smittade när en kontakt uppstår [23]. Detta kan vara lämpligt när det är ett begränsat samhälle man studerar, då mycket information behövs om samhällets struktur.

2.4 Mean Squared Error

Vi vill införa en kvantitativ metod för att bestämma hur bra modellerna är. Detta blir också ett sätt att jämföra de olika modellerna med varandra. Till ändamålet kommer vi att använda medelkvadratfel (MSE, *mean squared error*). Första steget blir då att dela upp tidsintervallet i m ekvidistanta delar. Vid varje tidssteg tar man kvadraten av skillnaden mellan datan och modellen,

vilket vi kallar $d_i^{data} - d_i^{modell}$, där d är någon typ av data som man vill att modellen följer. Då blir formeln för MSE

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (d_i^{data} - d_i^{modell})^2. \quad (27)$$

MSE är då ett mått på hur bra data följer modellen. Exakt vad man väljer att subtrahera för att få MSE tas upp i Metoden.

2.5 Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) är en grupp metoder som kan användas för att estimeras parametrar utan att behöva beräkna likelihoodfunktionen. Det grundar sig i Bayesansk statistik, där Bayes sats har en central roll

$$p(\Theta|D) = \frac{p(\Theta)p(D|\Theta)}{p(D)}. \quad (28)$$

Här är posteriorifördelningen $p(\theta|D)$ sannolikheten att en parameter får värdet θ givet data, $p(D|\theta)$ är likelihoodfördelningen som representerar sannolikheten att få data givet parametervärdet. Apriorifördelningen $p(\theta)$ är sannolikheten att erhålla ett visst parametervärde och $p(D)$ är sannolikheten att få datan. I vissa sammanhang kan likelihoodfördelningen erhållas analytiskt, men i andra fall är det mycket svårare. För bland annat individbaserade modeller som är väldigt komplexa är det ofta orimligt att ta fram likelihoodfördelningen [24]. I de fallen kan ABC användas till att uppskatta posteriorifördelningen. ABC grundar sig i att man utifrån en apriorifördelning väljer parametervärden som genererar dataset, vilka jämförs med observerad data genom att beräkna till exempel MSE. Om skattningen är tillräckligt bra sparas tillhörande parametervärden och används till att konstruera en posteriorifördelning [25].

3 Metod

I Teorin beskrev vi fyra olika modeller: den deterministiska SIR-modellen, Markovkedjemodellen, SDE:n och en individbaserad nätverksmodell. SIR-modellen bygger på kopplade differentialekvationer och saknar stokastiska element. Markovkedjemodellen där sjukdomen modelleras som en slumpvandring med två möjliga utfall. Markovprocessen kan sedan användas som en startpunkt för att formulera en stokastisk differentialekvation, i vilken de okorrelerade stegen i Markovprocessen kan beskrivas som en brusterm i form av en Wienerprocess. Slutligen så införde vi en individbaserad nätverksmodell där tillstånden inte är homogena utan behandlar varje individ som en nod i ett nätverk. Med dessa modeller och lämplig data beskrivs nu hur simuleringarna genomfördes.

3.1 Data

För att kunna jämföra och anpassa våra modeller till verkligheten har data från Socialstyrelsen[26] på antal personer som vårdats på sjukhus använts. Datan läggs ut veckovis och vid simulering använder vi senaste versionen, så att modellen ska ha så bra förutsättningar som möjligt att spegla verkligheten. Vissa modeller använder en annan tidsenhet än veckor, i dessa fall så använder vi en linjärt interpolerad version av datan. Icke-kumulativt antal användes vid anpassning till alla modeller förutom nätverksmodellen, där kumulativa antalet användes. En graf över antalet inläggningar som funktion av dagar kan ses i figur 11 i Appendix D.

3.1.1 Mobilitetsdata

Mobilitetsdatan som används är tagen från Googles mobilitetsstatistik [27]. Vad mobilitetsdatan beskriver är hur mycket människor rör sig i samhället. Den räknas ut med hjälp av platstjänstdatan från mobiler. Datan var given som procent från en viss vecka. Vi gjorde om den till ett mobilitetsindex som kan ses i figur 12 i Appendix D. Beräkningen såg ut såhär

$$\text{Mobilitetsindex}[i] = 1 + \text{mobilitetsdata}[i]/100. \quad (29)$$

Som i fallet med inläggningsdatan så använder vissa modeller en annan tidsenhet än veckor. I dessa fall så här vi använt en linjärt interpolerad version av datan.

3.2 Deterministisk SIR-modell

Simulationen av de deterministiska modellerna utfördes med Runge-Kutta-integrationsmetoden, vilken numeriskt löser differentialekvationer på formen: $\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x}(t))$. Metoden bygger på diskretisering av tillståndsvariablerna $\mathbf{x}(t) \rightarrow \mathbf{x}_k$ samt tidsintervallet. Sedan med hjälp av en så kallad "Butcher tableau" (som utgörs av a_{ij} och b_i), kan en generell Runge-Kutta metod av ordning s , uttryckas som

$$\begin{aligned} \mathbf{K}_i &= \mathbf{f} \left(\mathbf{x}_k + \Delta t \sum_{j=1}^s a_{ij} \mathbf{K}_j \right), \\ \mathbf{x}_{k+1} &= \mathbf{x}_k + \Delta t \sum_{i=1}^s b_i \mathbf{K}_i. \end{aligned} \tag{30}$$

I Python-paketet `scipy.integrate`, existerar funktionen `solve_ivp`, där ordningen kan specificeras. För detta arbete brukades den traditionella RK45 varianten. Detta är en hybrid mellan en ordning fyra och ordning fem metod som adaptivt kan ändra sin steglängd Δt .

3.2.1 SIR med mobilitetsdata

Modellen har också utökats med mobilitetsdata. Denna data definierar vi som $f_{mob.}(t)$ och kan ses i figur 12 i Appendix 3.1. Det är alltså den linjärt interpolerade versionen av Mobilitetsindex[i] som betecknas $f_{mob.}(t)$. Ett linjärt beroende infördes på β så att

$$\beta = \beta_1 + \beta_2 f_{mob.}(t). \tag{31}$$

Denna modell löstes sedan med Runge-Kutta. Anledningen till att man vill införa detta beroende är för att undersöka om detta kan leda till en modell som kan förutspå framtiden med hjälp av mobilitetsdatan.

Vi använde ABC för att få de bästa β , alltså de β som ger lägst MSE. Denna applikation av ABC antog inget om apriorifördelningen och därför användes en uniform fördelning för denna. Då simuleringen tog lång tid valde vi att begränsa oss till $\beta_1, \beta_2 \sim \mathcal{U}[0, 0, 3]$.

Sedan behöver vi en funktion för att bedöma om β skall sparas, vilken i vårt fall var MSE. I MSE:en definierades d_i som antalet inläggningar vid tiden i . För att veta hur många smittade som behöver sjukhusvård används `hospitalPar`, vilken vi beskriver mer om under sektion 3.5.2. som avgör hur många som behöver sjukhusvård, som här betecknas α . Med en maximal tolerans för MSE kommer vi att selektera värden på parametrarna som har MSE mindre än eller lika med toleransen. Om vi kallar denna tolerans MSE_{max} skulle sambandet vara

$$MSE = \frac{1}{t_{max}} \sum_{t=0}^{t_{max}} (\text{inläggningar}^{data}(t) - \alpha I^{modell}[t])^2 \leq MSE_{max} \tag{32}$$

där (t) betecknar värde på tiden och $[t]$ betecknar index för fältet med tider. Denna process upprepas sedan ett antal gånger. Hur många simulationer som behövs beror hur stor andel som klarar av toleransen.

3.3 Markovkedjemodell

Simulationen gjordes i Python med Gillespies algoritim. För beskrivning av denna algoritim, se Appendix B.5. I vårt fall användes u_1 för att beskriva vid vilken tid händelsen sker. En slumpmässigt

vald tid från T betecknas som τ . Då är

$$\tau = \frac{\ln u_1}{\lambda}. \quad (33)$$

För en härledning av denna tid, se Appendix B.6. Bestämning av vilken händelse som sker avgörs med variabeln u_2 . Först delades det likformiga intervallet så att varje händelse har en motsvarande stor del av intervallet. I vårt fall så har vi två händelser och vi behöver då dela upp intervallet $[0, 1]$ i två delintervall, $[0, p_1]$ och $(p_1, 1]$, där p_1 är sannolikheten är smitta och kommer beskrivas nedan. Om $u_2 \in [0, p_1]$ så blir $(k, j) = (-1, +1)$ annars $(k, j) = (0, -1)$ [7]. Detta fortsätter tills att $i = 0$ eller $t > t_{max}$ där t_{max} är någon maximal tid för simulationen.

Sannolikheten att gå från ett tillstånd i till j definieras som $p_{ij} = \frac{q_{ij}}{q_i}$ [6]. Sannolikheten att gå till $(k, j) = (-1, +1)$, alltså smitta, definierar vi som p_1 , vilken uttrycks som

$$p_1 := \frac{Q_{(k,j)=(-1,+1)}}{\sum Q} = \frac{\beta i \frac{s}{N}}{\beta i \frac{s}{N} + \gamma i}. \quad (34)$$

På samma sätt är sannolikheten att gå till $(k, j) = (0, -1)$, återhämtning, definierad som

$$p_2 := \frac{Q_{(k,j)=(0,-1)}}{\sum Q} = \frac{\gamma i}{\beta i \frac{s}{N} + \gamma i}. \quad (35)$$

Då det endast existerar två möjliga tillstånd innebär detta att relationen mellan sannolikheterna trivialt kan uttryckas som $p_2 = 1 - p_1$.

Utöver övergångssannolikheterna behövs även tiden mellan händelserna. Här ger Markovegenskapen att denna tid T är exponentialfördelad som $T \sim \lambda e^{-\lambda t}$, där λ är summan av alla övergångstakter för alla möjliga händelser [7]. För SIR-modell med Markovkedjor i kontinuerlig tid,

$$\lambda = \beta i \frac{s}{N} + \gamma i \quad (36)$$

där (s, i) är värdet för $(S(t), I(t))$ evaluerat vid en tidpunkt t .

Vi använde ABC för att få de bästa β , alltså de β som ger lägst MSE. Denna applikation av ABC antog inget om apriorifördelningen och därför användes en uniform fördelning, $\beta \sim \mathcal{U}[0, 1]$, för denna. Sedan behöver vi en funktion för att bedöma om β skall sparas, vilken i vårt fall var MSE. En skillnad från fallet SIR med mobilitetsdata, är att Markovmodellen inte klarar av höga N . Detta eftersom en fördubbling av antalet personer kräver två extra simuleringar av både u_1 och u_2 . Tidkomplexiteten är alltså $\mathcal{O}(N)$. För att lösa detta skalar vi MSE med en faktor κ så att

$$\text{MSE} = \frac{1}{t_{max}} \sum_{t=0}^{t_{max}} (\kappa \cdot \text{inläggningar}^{data}(t) - \alpha \cdot I^{modell}(t))^2 \leq \text{MSE}_{max} \quad (37)$$

där κ avgör förhållandet mellan N_{modell} i modellen och N_{data} i datan (Sveriges befolkning). Lagg också märke till att vi använde $I^{modell}(t)$, inte $I^{modell}[t]$. Detta eftersom avstånden mellan I är olika varje gång och vi därför också gjort en linjär interpolation för I^{modell} . Skalningen leder dock till att MSE inte kan jämföras direkt mellan modellerna. Hur många gånger som modellen behöver simuleras beror på hur stor andel som klarar av toleransen. Annars är processen likadan för Markovkedjemodellen som för SIR med mobilitetsdata.

Sedan, med konstant β , simulerades sannolikheten att spridningen slutar i tidigt skede. Tidigt skede definierades som att antalet smittade vid tid 100 måste vara större än 1. Sannolikheten att spridningen slutar i tidigt skede simulerades med ett varierande antal infekterade vid tid noll, I_0 .

3.4 SDE SIR-modell

Simulationen av SDE-modellen har genomförts i Python. För detta användes Euler-Maruyama-metoden. Detta är ett sätt att få numeriska lösningar till stokastiska differentialekvationer. Vad

metoden innebär i praktiken är att vi byter ut de kontinuerliga delarna mot numeriska approximationer. Vi diskretiserade tiden med tidsteg Δt och formulerade ekvationerna som

$$\begin{aligned}
 dS(t) &\approx S(t + \Delta t) - S(t) = \\
 &= -\frac{\beta S(t)I(t)}{N} \Delta t - \sqrt{\frac{\beta S(t)I(t)}{N(t)}} \underbrace{(W_1(t + \Delta t) - W_1(t))}_{\sim \eta_1 \sqrt{\Delta t}} \quad (\text{Kvadratisk variation}) \\
 \Leftrightarrow & \boxed{S(t + \Delta t) = S(t) - \frac{\beta S(t)I(t)}{N} \Delta t - \sqrt{\frac{\beta S(t)I(t)}{N(t)}} \eta_1 \sqrt{\Delta t}}. \quad (38)
 \end{aligned}$$

På samma sätt följer:

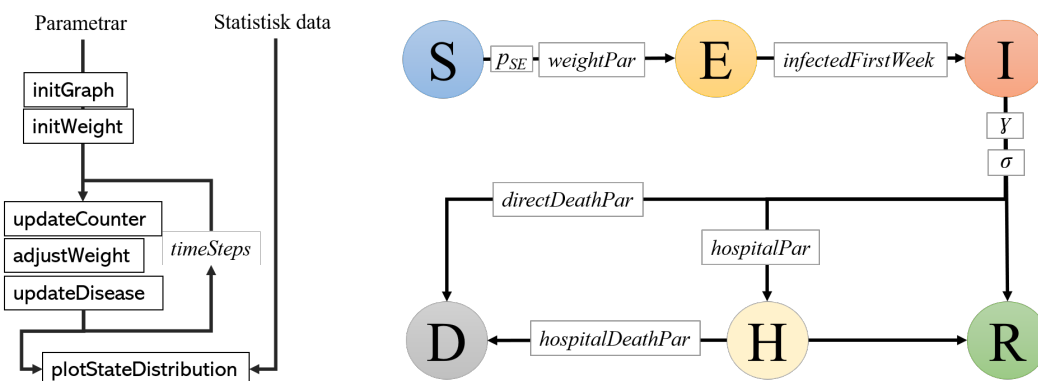
$$\boxed{I(t + \Delta t) = I(t) + \left(\frac{\beta S(t)I(t)}{N} - \gamma I(t) \right) \Delta t + \sqrt{\frac{\beta S(t)I(t)}{N}} \eta_1 \sqrt{\Delta t} - \sqrt{\gamma I(t)} \eta_2 \sqrt{\Delta t}}.$$

där η_1, η_2 är oberoende standard-normalfördelade variabler. Parametrar skattades med hjälp av ABC för att få de β som ger lägst MSE. Detaljerna är densamma som vid simuleringen av SIR med mobilitetsdata.

Sedan, med konstant β , simulerades sannolikheten att spridningen slutar i tidigt skede. Sannolikheten att spridningen slutar i tidigt skede simulerades med ett varierande antal infekterade vid tid noll, I_0 . Detaljerna är här densamma som vid simuleringen av Markovkedjemodellen.

3.5 Nätverksmodell

Den utvecklade nätverksmodellen har diskreta tillstånd enligt en SEIR-modell med ytterligare tillstånd för avlidna (tillståndet benämns D) och de som blir vårdade på sjukhus (H). Det är en individbaserad modell där individer rör sig mellan de olika tillstånden. Modellen är programmerad i Python och koden finns på GitHub, se Appendix E. Här kommer vi nu gå genom modellen utifrån de funktioner som körs i programmet. Figur 3 till vänster visar hur programmet är uppbyggt: ett antal parametrar används till att initiera grafen, varpå ett antal tidsteg följer där individer smittas, blir sjuka, friska och så vidare, beroende på andra parametrar. Sedan plottas antal individer i varje tillstånd för varje tidsteg tillsammans med statistisk data för att kunna jämföra och anpassa modellen till verkligheten. Datan som används är enligt avsnitt 3.1 det kumulativa antalet sjukhusinläggningar i Sverige.



Figur 3: Till vänster: Schema över hur programmet är uppbyggt för nätverksmodellen. Statistisk data är kumulativt antal sjukhusinläggningar i Sverige. Till höger: Flödesschema över hur individer rör sig mellan de olika tillstånden i nätverksmodellen. Parametrar som anger villkor för varje övergång finns även ritade.

Den första funktionen som används i modellen är `initGraph`. Denna funktion genererar modellens grundtillstånd (innan smittspridningen börjat), där modulen `networkx` används för att bygga upp

det sociala nätverket som simuleringen körs på. Funktionen `nx.barabasi_albert` tar två parametrar, antalet noder `numberOfNodes` och antal kanter som ska bildas till varje ny nod m . Parametern `numberOfNodes` sätts till 10000 för att en simulering inte ska ta för lång tid och m är 2 enligt en tidigare studie [28]. Alla noder börjar med att tillhöra kategorin S genom att ange det som ett attribut till varje nod. Det sista som sker i `initGraph` är att en procentsats på 0,5 % av det totala antalet noder blir smittade och flyttas till kategori I. Dessutom blir den nod som har 50:onde flest kopplingar också sjuk, detta för att endast studera då smittspridningen har påbörjats. Vi undersöker alltså inte körningar som resulterar i att smittan dör ut i ett tidigt skede.

Efter grafen har initierats kommer vikter initieras i `initWeight`. Varje kant i nätverket får vikten 1 (ett), detta representerar grundtillståndet i populationen då man träffar alla personer i sin umgängeskrets som vanligt. Dessa vikter kommer ändras i `adjustWeight` för att modellera hur social distancering och andra åtgärder påverkar smittspridningen. Villkoret som måste uppfyllas för att vikterna ska ändras är att smittspridningen har uppnått en viss procentsats av befolkningen. I koden används 1% som gräns, vilket är ungefär där smittspridningen låg efter den första vågen av Covid-19 2020. Detta är dock en mycket grov uppskattning då testningen av personer var undermålig första halvåret. Vikterna minskar med en viss andel enligt `weightPar`, vilken estimeras efter data. När smittan sedan gått ned under 0,5% ökas vikterna igen till grundtillståndet.

I `updateCounter` analyseras varje individs tillstånd och sedan lagras antalet tillhörande i varje tillstånd. Detta sker för varje tidsteg så att det är lätt att få en överblick av smittans spridning över tid.

Den simulerade smittspridningen sker i `updateDisease`, där individer flyttas mellan tillstånden i figur 3 till höger. Observera att pilarna endast har en riktning och det är inte möjligt att bli smittad mer än en gång. Individer rör sig mellan de olika tillstånden enligt ett antal parametrar, dessa har uppskattats genom litteraturstudier och anpassning till data, vilket beskrivs i avsnitt 3.5.1 och 3.5.2.

I koden avgörs varje flytt av en nod från ett tillstånd till ett annat av förbestämda regler. En nod kan flyttas från S till E om den träffar på en infekterad nod och blir smittad. Att träffa på en infekterad nod bestäms av om noden har en kant till någon infekterad och om vikten är tillräckligt hög, vilket påverkas av `weightPar`. Att bli smittad bestäms av parametern p_{SE} som anger sannolikheten att bli smittad vid en interaktion med en smittad individ. Detta liknar parametern som tillsammans med antal kontakter per tid k bildar β i SIR-modellen, vilket vi kort nämnde i Teorin. Skillnaden är att p i SIR-modellen beskriver sannolikhet att bli smittad vid kontakt med en godtycklig individ.

Inkubationsperioden i E modelleras genom att en andel av noderna som kommer dit är där i en vecka, där andelen ges av parametern `infectedFirstWeek`, och om noden inte flyttats till I efter första veckan kommer den flyttas dit vid nästa veckostegning och då kunna infektera andra noder.

En utväg från att vara infekterad är att hamna på sjukhus. Det sker om noden varit smittad tillräckligt länge, vilket beror på parametern γ och dess standardavvikelse σ . Sedan inför vi parametern `hospitalPar` som avgör andelen av de smittade som behöver sjukhusvård. En annan utväg från att vara infekterad är att dö, vilket parametern `directDeathPar` avgör. Det sker endast om noden varit infekterat tillräckligt länge, alltså påverkar γ och σ utfallet. Den stora majoriteten av befolkningen blir sjuk och tillfrisknar sedan. I koden kan en nod flyttas direkt till R om noden varit i I tillräckligt länge och den inte dör eller hamnar på sjukhus. Utgången från att bli inlagd på sjukhus är antingen att bli friskförklarad och det bestäms av `1-hospitalDeathPar`, eller så hamnar noden i D, vilket då sker beroende på `hospitalDeathPar`.

I `plotStateDistribution` visualiseras sjukdomens spridning, samt dess konsekvenser genom att använda den lagrade datan från `updateCounter` för att skapa grafer med antalet friska, sjuka, återhämtade, avlidna och sjukhusinlagda per vecka. Dessa grafer skapas med hjälp av modulerna `matplotlib` och `seaborn`, som är verktyg till Python för att skapa och anpassa grafer.

3.5.1 Beräkningsmetod för parameterskattning

De olika parametrar vi presenterat ovan i avsnitt 3.5 uppskattas med litteraturstudier och med beräkningsmetoder. De parametrar som bestäms från litteraturen är *incubationTime*, *gamma*, *hospitalPar*, *hospitalDeathPar* och *directDeathPar*. Alltså återstår *p_{SE}* och *weightPar* att bestämma med beräkningsmetoder.

ABC (Approximate Bayesian Computation), som vi presenterade i Teorin, används. För ABC väljer vi att använda en uniform apriorifördelning mellan noll och ett, för att vi inte vet något om de parametrar som ska skattas. Efter några initiala testkörningar justeras storleken på intervallet till att fånga in de intressanta värdena och inte slösa körningar på värden som är orimliga. Sedan används MSE för att jämföra data från modellen med observerad data. Villkoret för acceptans väljs till att en körningen av modellen ska resultera i en skattning av sjukhusinläggningar som är tre gånger bättre än att inga inläggningar sker, alltså att antalet ligger konstant på noll.

3.5.2 Parameterskattning från data

Fördelningen för tiden mellan att en person smittats till att symptom uppträder, baseras på en studie [29] gjord i början av pandemin. Studien omfattar 181 fall mellan den 4 januari och 24 februari 2020, där en majoritet av fallen var invånare av eller personer som nyligen rest till den kinesiska provinsen Wuhan. Fördelningen beskrivs som en lognormalfördelning, där 78% visade symptom inom en vecka och för majoriteten av resten dröjde det två veckor innan symptom uppvisades. Endast 1% hade en inkubationstid längre än två veckor och vi har valt att räkna in dem i andelen som visar symptom under andra veckan i koden.

Takten för hur en infekterad person tillfrisknar har i en studie[30] beskrivits som en normalfördelning, som vi har applicerat i vår kod. Studien är en sammanställning av 5769 israeliska fall publicerade av the Isreal Ministry of Health den 28 april. Datan är uppdelad i ålder och kön, men då vår modell inte tar hänsyn till dessa faktorer beräknades medelvärdet som kan appliceras på hela befolkningen.

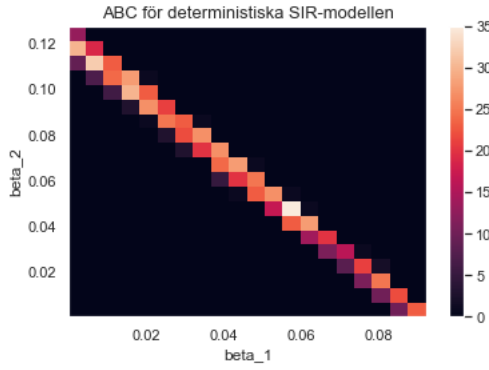
Parametern *hospitalDeathPar* uppskattades utifrån en svensk kohortstudie som baserar sig på sjukhusinläggningar nationellt under tidsperioden mars till september under 2020 [31]. Då totalt 14755 patienter vårdades inom slutenvården (IVA exkluderat) under denna period varav 2339 omkom uppskattades denna parameter till 15,9%. Vidare användes en rapport av Folkhälsomyndigheten[32] för att skatta parametern *directDeathPar* som är en uppskattning dödligheten över hela samhället orsakad av Covid-19, enligt Folkhälsomyndighetens undersökning, som baserar sig på dödsfall i Stockholms-regionen uppskattas infektionsdödligheten vara 0,6%. Till sist uppskattades *hospitalPar* till 5%, vilket gjordes genom att använda sig av Socialstyrelsens statistik på antalet slutenvårdade Covid-19-patienter[26] tillsammans med Folkhälsomyndighetens veckorapport[33] om det totala antalet bekräftade fall av Covid-19 i Sverige. För mer detaljerad beskrivning, se Appendix F.

4 Resultat

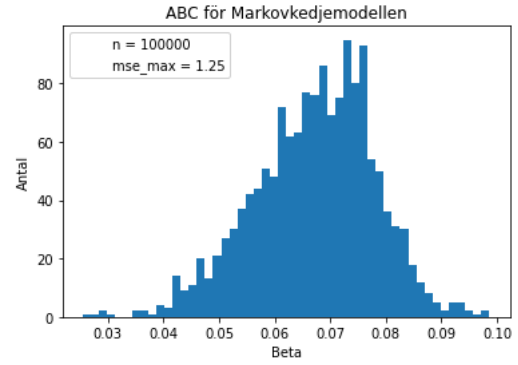
I detta stycke presenteras ett antal olika resultat som är till följd av bland annat de beräkningsmetoder som använts för att optimera de olika modellerna. Detta kommer användas för att jämföra de framtagna modellerna, men även för att pröva deras validitet.

4.1 SIR med mobilitetsdata

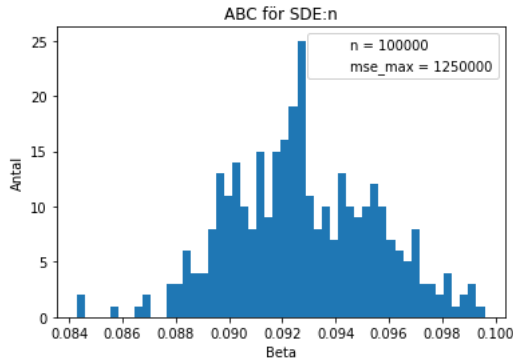
ABC-simuleringen har genomförts med $N = 10^7$, $\gamma = 1/14$, $hospitalPar = 0,05$, $I_0 = 10$. Apriorifördelningen var $\beta_1, \beta_2 \sim \mathcal{U}[0, 0, 3]$. Resultatet av $n = 10^5$ simuleringar med $MSE_{max} = 1,25 \times 10^6$ som maximal tolerans kan ses i figur 4a.



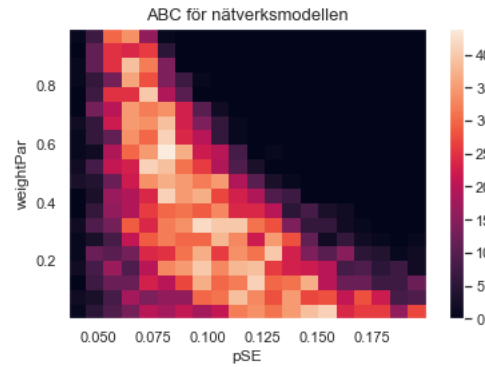
(a) ABC-simulering för SIR-modellen med mobilitetsdata med $N = 10^7$, $n = 10^5$, $\gamma = 1/14$, $hospitalPar = 0,05$, $I_0 = 10$.



(b) ABC-simulering för Markovkedjemodellen med $N = 10^4$, $n = 10^5$, $\gamma = 1/14$, $hospitalPar = 0,05$, $I_0 = 10$, $\kappa = 10^{-3}$, $t_{max} = 500$.



(c) ABC-simulering för SDE:en med $N = 10^7$, $n = 10^5$, $\gamma = 1/14$, $hospitalPar = 0,05$, $I_0 = 10$, $T = 385$, $dt = 1$.



(d) ABC-simulering av nätverksmodellen med $n = 1,9 \times 10^4$ körningar av modellen.

Figur 4: Resultat av ABC för de fyra olika modellerna

4.2 Anpassning av stokastiska modeller

De gemensamma parametrarna var $\gamma = 1/14$, $hospitalPar = 0,05$, $I_0 = 10$. Apriorifördelningen var $\beta \sim \mathcal{U}[0,01, 1]$. Anpassningen gjordes på hela datan.

4.2.1 Markovkedjemodellen

Resultatet av n simuleringar med MSE_{max} som maximal tolerans, en maximal tid $t_{max} = 500$ samt en skalningsfaktor på $\kappa = 10^{-3}$ kan ses i figur 4b. Den skalade populationen var $N_{Markov} = N\kappa = 10^4$.

4.2.2 SDE-modellen

Resultatet av n simuleringar med viss MSE_{max} som maximal tolerans, en total tid på $T = 385$ och en tidssteg på $dt = 1$ kan ses i figur 4c. Vi använde $T = 385$ eftersom datan går till 385.

4.2.3 Sannolikhet att pandemin dör ut

Resultatet av $n = 10^4$ simuleringar där vi lägger till ett till en räknare varje gång som pandemin dör tidigt. Parametrarna var likdana som tidigare simulationer fast med konstanta $\beta_{Markov} = 0,075$ och $\beta_{SDE} = 0,09$. Resultatet kan ses i figur 14a under Appendix G. I denna figur kan vi se sannolikheten att pandemin dör minskar med avseende på I_0 , antalet smittade vid tid noll. Skillnaden i proportionerna av döda pandemier mellan Markovkedjemodellen och SDE:n kommer från skillnaden i \mathcal{R}_0 . I figur 14b kan vi se sannolikheterna ifall samma parametrar användes för modellerna.

4.3 Spridning för SDE och Markovmodellen

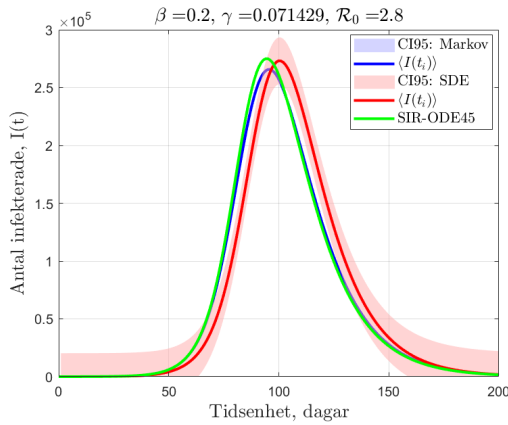
Variationen för SDE:n och Markov modellen ritades med respektive medelrealiseringar och numeriska lösningen till SIR med $N = 10^6$, $I(0) = 10$, $S(0) = N - I(0)$. På grund av minnesbegränsningar kördes endast $n = 100$ simulationer för Markovmodellen, medan SDE:n enkelt kunde simuleras $n = 10^6$ gånger. Medelinstanserna för en tidpunkt t_i beräknades på följande sätt

$$\langle I(t_i) \rangle = \frac{1}{n} \sum_{j=1}^n I^{(j)}(t_i), \quad 0 \leq t_i \leq T_{max}. \quad (39)$$

För Markov-modellen krävdes att datan interpolerades, som är en konsekvens av slumpprocessen för tidstegen mellan utfall. Slutligen, eftersom SDE:n producerade värden för varje tidpunkt krävdes ingen extra behandling av datan. Konfidensintervallen för båda modellerna beräknades genom att använda σ_{max} . Standardavvikelsen beräknades med medel antalet infekterade som referens, därmed erhöles flera olika värden för standardavvikelsen för varje simulering. Slutligen producerades konfidensintervallet med t-fördelningsvärdet för konfidensgraden 95% med m observationer för varje simulering

$$CI_{95} = \left(\langle I(t) \rangle - t_{95} \frac{\sigma_{max}}{\sqrt{m}}, \langle I(t) \rangle + t_{95} \frac{\sigma_{max}}{\sqrt{m}} \right), \quad m = \lfloor [0, T_{max}] \rfloor. \quad (40)$$

I figur 5 är antalet individer inom I -tillståndet illustrerat.



Figur 5: Spridningen av medelrealiseringar samt 95% konfidensintervall för dem. SIR-modellens numeriska lösning är även inkluderad för att uppvisa hur medelrealiseringarna närmar sig samma lösning. Figuren påvisar att spridningen var större för Itô-SDE:n än för Markovmodellen.

I figur 5 kan det ses att konfidensintervallet på Itô-drift ekvationen var bredare än för Markovmodellen. I regel varierade även SDE:n mer än Markov-modellen. I Appendix H, finns figurerna 15a,15b,15c där variationen för några olika populations storlekar samt \mathcal{R}_0 -värden för Markovmodellen urskiljas. Figurerna påvisar att \mathcal{R}_0 -parametern och populationsstorlek påverkade variansen för Markovprocessen. En liknande spridning uppstod även för SDE modellen, se figurerna 16a,16b,16c i Appendix I.

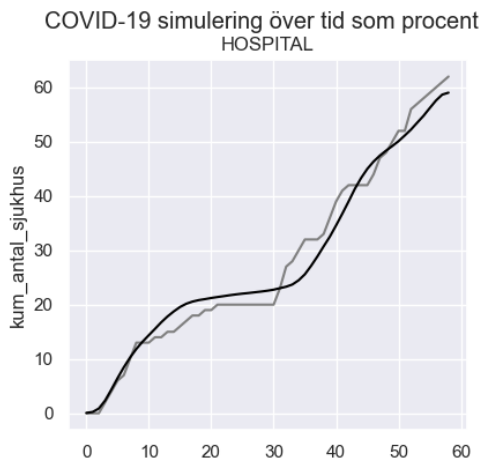
4.4 Anpassning av nätverksmodell

Nätverksmodellen anpassas till antal sjukhusinläggningar som vi beskrivit i Metoden i avsnitt 3.5.2. Utöver de parametrar som erhållits från litteraturstudier återstår parametrarna p_{SE} och $weightPar$. Modellen simulerades med en ABC-algoritm med apriorifördelning $p_{SE} \sim \mathcal{U}[0, 0, 2]$ och $weightPar \sim \mathcal{U}[0, 1]$. Modellen simulerades $n = 1,9 \times 10^4$ gånger, vilket resulterade i grafen i figur 4d.

4.5 Möjligt resultat från nätverksmodell

Vid 100 stycken simuleringar av nätverksmodellen där parametrarna från litteratur användes, samt en kombination av $p_{SE} = 0,081$ och $weightPar = 0,43$ som togs från resultatet av ABC-simulering i figur 4d erhöles följande resultat som det bästa. De olika simuleringarna jämfördes med hjälp av R^2 -talet som är ett mått på variansen mellan de beroende (Simuleringen) och de oberoende

(Sjukhusdatan) mätvärden. Det bästa erhållna resultatet hade ett R^2 -tal som ungefär var $\approx 0,97$, ett $R^2 = 1$ hade indikerat en perfekt anpassningen. Se figur 6 nedan för grafisk visualisering av resultatet.

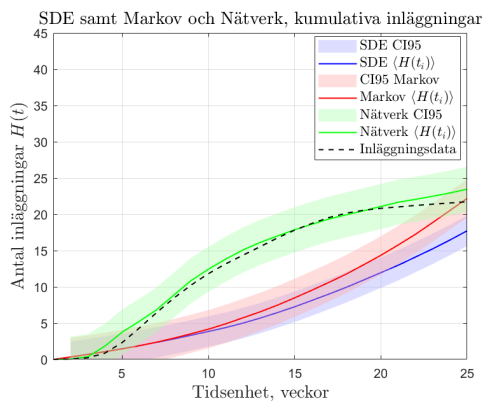


Figur 6: Bästa erhållna resultat från 100 simuleringar, $R^2 \approx 0.973$. Den svarta linjen är data från sjukhusinläggningar och den grå linjen är en simulering av nätverksmodellen.

4.6 Kumulativa inläggningar för SDE-, Markov- och nätverksmodellen

Alla modeller simulerades $n = 100$ gånger, medan populationen var $N = 10^4$. För SDE:n och Markovmodellen användes β -värden som maximerade respektive parameterfördelning för nya ABC simuleringar med tidigare nämnda parametrar. Vid denna simulation användes data från endast de 25 första veckorna. Interaktionsparametrarna blev därför $\beta_{SDE} = 0,082$, $\beta_{Markov} = 0,085$. Resten av parametrarna och begynnelsevilkoren var samma, $\gamma = 1/14$, $I(0) = 10$ och $hospitalPar = 0,05$. För nätverksmodellen var $p_{SE} = 0,07$ och $weightPar = 0,5$. En figur som redovisar samtliga stokastiska modeller anpassade till inläggningsdata kan skådas i figur 7.

Alltså var $\mathcal{R}_0^{Markov} = 1,190$ och $\mathcal{R}_0^{SDE} = 1,148$.



Figur 7: I figuren är det kumulativa antalet inläggningar för Markovmodellen, Itô-SDE:n och Nätverksmodellen illustrerat. Notera att för Markovmodellen och Itô-SDE:n beräknades inläggningar med: $\langle H(t_i) \rangle = hospitalPar \cdot [S(\langle S(t_{i+1}) \rangle) - \langle S(t_i) \rangle]$, där $t_{i+1} - t_i = 1$ vecka. De stokastiska modellerna uppnår ungefär samma värde vid sluttiden, dock följer deras grafer inte datan lika väl som nätverksmodellen. Detta beror på det exponentiella beteendet för de stokastiska modellerna, alltså är detta en begränsning för dessa modeller.

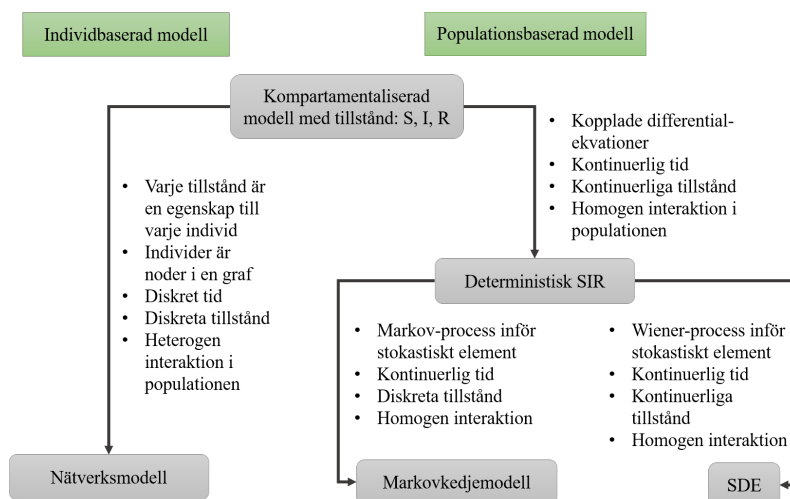
Respektive modell hade totala felet i ökande ordning, $\varepsilon = \sum_i |H(t_i) - H_{data}(t_i)|$: $\varepsilon_{Nät} = 14,347$, $\varepsilon_{Markov} = 126,472$, $\varepsilon_{SDE} = 157,991$.

5 Diskussion

Med utgångspunkt i den deterministiska SIR-modellen som Kermack och McKendrick publicerade för nästan ett sekel sedan har vi utvecklat fyra andra modeller för att beskriva smittspridningen av Covid-19. En deterministisk SIR-modell, två stokastiska SIR-modeller och en nätverksmodell. Var och en har sina styrkor och svagheter, och beroende på vilken fråga man ställer kommer det variera vilken modell som kan ge bäst svar.

5.1 Jämförelse av de framtagna modellerna

De olika modellerna kan delas upp i tre populationsbaserade modeller och en individbaserad modell. De populationsbaserade modellerna beskriver smittspridningen på populationsnivå och befolkningen delas upp i olika tillstånd, till skillnad från den individbaserade där varje individ kan vara i varje tillstånd. Övergången från populationsbaserad till individbaserad modellering ligger i att man går från att studera flöden av individer till att studera den faktiska interaktionen mellan individer som ger upphov till smittspridning. Figur 8 beskriver hur alla modeller hänger ihop och de antaganden som görs till varje modell.



Figur 8: Schematisk bild över de olika modellerna och hur de hänger ihop.

Med hjälp av nätverksmodellen skulle man kunna underlätta smittspårning då noderna i modellen representerar enskilda individer, vilket är något som inte är möjligt med de populationsbaserade modellerna. Anledning till att den lämpar sig bättre än t.ex. Markovkedjemodellen är att sannolikheten att bli sjuk beror på hela befolkningen, men i nätverksmodellen beror det på proximitet. Man kan alltså studera mekanismerna bakom smittspridning.

Gällande precision kommer det för de stokastiska modellerna variera för varje körning, men den som har gett lägst MSE för hela datamängden är nätverksmodellen. De andra modellerna har inte i dagsläget möjlighet till att beskriva en andra våg. Om endast en våg studeras är det även nätverksmodellen som ger bäst anpassning. Den deterministiska SIR-modellen ger alltid samma lösning för samma parameterkombination och kan därför anses vara mest robust.

Tiden för en simulering kan vara intressant om många simuleringar ska köras. Den deterministiska SIR-modellen är snabbast med Runge-Kutta-metoden och sedan kommer SDE:n, följt av Markovkedjan och långsammast är nätverksmodellen. Det kan alltså i det avseendet vara mer intressant att använda en av de snabbare modellerna när man vill köra många simuleringar.

Med många parameter kommer också stor osäkerhet, vilket kommer från antaganden man gör vid estimeringen av parametrarna. Nätverksmodellen har fler parametrar än vad den deterministiska och de två stokastiska SIR-modellerna har, och därtill även en stor osäkerhet. Det man vinner i att beskriva smittspridningen mer detaljerat förlorar man alltså i säkerhet. Dock kan den mer detaljerade beskrivningen ge en mer precis beskrivning av verkligheten.

5.2 SIR med mobilitetsdata

Från figur 4a kan vi se att summan $\beta_1 + \beta_2$ är linjär. Det är faktiskt nästan så att denna summa är konstant. Detta pekar på att mobilitetsdata inte tillför något till hur bra modellen kan anpassa sig till datan. Anledningen till att vi får detta resultat kan vi bara spekulera i. Antagligen får vi detta eftersom SIR är en lite för enkel modell. Linjärt beroende på β räcker kanske inte för att

få flera vågor. Andra modeller hade kanske presterat bättre. Till exempel en SIS-modell eller en SIR-modell med *breakpoints*, punkter där man låter β ändras.

5.3 Itô SDE:n och Markovmodellen

För Itô SDE:n dominerar drift-termen $\mu(t, \mathbf{X})$ för majoriteten av simulationerna. Detta medför att instanser som går mot en exponentiell ökningsfas kommer att domineras av det underliggande SIR-modell ramverket. Efter 10^6 simuleringar, var den medelvärdesbildade instansen $\langle I(t_i) \rangle$ ganska lik resultatet från Runge-Kutta lösningen av SIR-modellen vilket kan ses i figur 5.

Situationer där smittspridningen inte blir omfattande och dör ut kan endast ske om den stokastiska termen $G_t W_t$ dominerade över driften. Mer allmänt har det även konstaterats tidigare att SDE:er som härleddes från deterministiska differentialekvationer som SIR-modellen i allmänhet inte bevarar egenskapen att tillstånden ej kan anta negativa värden. Detta är därmed en begränsning för dessa modeller, i [34] presenterar Cresson och Sonner en proposition om det förgående påståendet, samt påpekar i synnerhet att en epidemisk modell (i det fallet en SIS-modell) inte bevarar endast positiva lösningar som garanteras av den ordinarie differentialekvationen. Detta är dock endast en detalj som är intressant ur ett teoriperspektiv då en numerisk simulation kan programmeras så att värdet för I blir noll om det blir negativt.

Markovmodellen likt SDE:n kommer även för stora populationer konvergera mot SIR-lösningen vilket också framträdde ur simulationerna, se figur 5. Anledningen för detta redogjordes främst under Teorisektionen för Itô-SDE:n. Intressant nog var spridningen mindre för Markovmodellen än för Itô-SDE:n vilket kan ses i Appendix H och I genom att jämföra figurerna, 15a-15c och 16a-16c. Den naturliga variationen medför därför att optimala parametrar endast kan uttryckas som fördelningar och inte som skalärer. Fördelningarna av β -parametern ser något annorlunda ut för SDE-modellen och Markovmodellen, se figur 4. Vilket styrks av den tidigare påpekade spridningen som uppträdde för både Markovmodellen och SDE:n. När de båda modellerna jämfördes mot verklig inläggningsdata var modellernas SIR-ramverk en tydlig begränsning, se figur 7. Det exponentiella beteendet båda modellerna uppvisar, är en allt för grov förenkling i fallet för Covid-19. Modellerna är dock enklare ur en simulerings synpunkt, där SDE-modellen är enklast att simulera följt av Markov-modellen.

5.4 Nätverksmodellen

Nätverksmodellen erbjuder många möjligheter till att härma beteende i befolkningen och struktur i samhället. Något att bygga vidare på är att mobilitetsdata skulle kunna användas på samma sätt som för den deterministiska SIR-modellen, men de försök vi gjort hittills har inte varit framgångsrika. En annan sak som kan studeras är att ta hänsyn till säsongeffekten, att det är vanligare att man är ute under sommarhalvåret och tvärt om. Dessutom kan man införa åldersgrupper, vilket många studier påvisat[30][35] spelar stor roll i mottagligheten och dödligheten hos viruset. Dock kräver en full implementering av åldersberoende i modellen känd data på hur åldern av en individ påverkar faktorer såsom kontaktnät, dödlighet, smittorisk vid kontakt med en infekterad, inkuberingstid och andra parametervärden. Detta är bara några exempel på hur man kan jobba vidare och det finns väldigt stor potential till att utöka modellen. Dock kommer ytterligare tillägg även innebära fler antaganden.

Ett av de mest omfattande antagandena vi gjort i dataanpassningen är att använda data från hela befolkningen i Sverige och sedan skala ner den för att kunna jämföra med storleken på vår simulering. Storleken på nätverken är 10000 noder, för att varje körning inte ska ta för lång tid. Grundtanken med individbaserade modeller är just att varje individ representerar en person i verkligheten, men vi har alltså gjort en medelbild av hela befolkningen. Då befolkningen blir tillräckligt stor spelar inte varje individs beteende lika stor roll, utan allt klumpas ihop och det som blir intressant är mönstren i beteende. Det är mönstren som vi försökt efterlikna i vår modell, men som sagt är detta en stor förenkling och huruvida detta antagande är rimligt bör undersökas vidare på samma och andra datamängder.

Vidare skulle uppskattningen av parametrar kunna förbättras då vissa parametrar uppskattats utifrån rapporter och litteratur. Till exempel uppskattas andelen personer som behöver sjukhusvård, *hospitalPar*, med hjälp av rapporter på antal som fått slutenvård och antal diagnostiserade med Covid-19. Båda dessa mängder är eventuellt underskattade, detta då slutenvårdsdatan bara inkluderar patienter man kunnat följa upp efter 30-dagar, vilket inte nödvändigtvis är alla. Det totala antalet sjuka i Covid-19 är med största sannolikhet en underskattning, dels då man inte testade lika många i början av pandemin, och dels för att många patienter varit asymptomatiska. Gällande andra parametrar som inkubationstiden och hur länge man är sjuk, kommer datan från rapporter skrivna i tidigare skeden av pandemin eller utförda i en annan del av världen. När dessa parametrar sedan appliceras på den svenska smittospridningen finns det risk att en viss avvikelse uppstår.

5.5 Samhälleliga och etiska aspekter

Projektbeskrivningen till detta kandidatarbete formulerades i samband med den globala pandemin orsakad av Covid-19. Målet var att skapa modeller som kan beskriva smittans utsträckning så här långt i samhället med visioner om att potentiellt kunna förutspå pandemins fortsatta spridning. Modellerna och deras applikationer kan användas till framtida forskning om virusets spridning. Till exempel kan modellerna användas till att förutspå eventuell sjukhusbrist eller studera effekten av olika restriktioner.

Det finns inga direkta etiska problem med hur vi genomfört projektet eller det resultat som vi kommit fram till. Låt oss därför anta att de modeller vi kommit fram till skulle tillämpas direkt av Folkhälsomyndigheten, då tillkommer stora följor på samhället. Om modellerna inte speglar verkligheten och felaktiga beslut tas skulle ett sådant beslut indirekt komma från detta projekt. En möjlighet är att modellerna felaktigt indikerar att nedstängning av samhället är det enda som kan göras för att hindra ytterligare spridning. Att stänga ned samhället påverkar många personers mentala och fysiska hälsa, vilket måste vägas mot den möjliga effekten av smittspridning. Om modellerna skulle visa att risken för stor spridning i samhället är mindre än vad den egentligen är kommer beslutsfattare inte vidta lämpliga åtgärder och individer kan smittas i högre utsträckning.

Införandet av individbaserade modeller kan göra det lättare att förstå hur en persons handlingar kan påverka ett helt samhälle och har möjlighet till att öka medvetenheten i samhället om att var individs handlingar har betydelse i den större situationen. Exempelvis kan en framtida utveckling vara att göra ett interaktivt program där man själv som privatperson kan se effekten av olika handlingar och på så sätt få större förståelse.

5.6 Avslutning

I rapporten har fyra olika modeller studerats, alla modeller använde olika fack för att klassificera varje individ i populationen. SIR-modellen var en utgångspunkt som förgrenades ut i två olika strategier, en mer överskådlig helhetsbild av populationen samt en individbaserad. Den överskådliga strategin innebär att populationen var homogen med ett slumpartat kollisionsbeteende. Dessa modeller är ganska bra på att beskriva luftburna sjukdomar som enkelt sprids i populationen. Variationen som uppträdde i körningarna för Markovmodellen och den stokastiska differentialekvationsmodellen (SDE:n) var ganska olika och i regel varierade Markovmodellen mindre för samma parametrar. Baserat på en jämförelse med inläggningsdata visade det sig att förenklingarna i de populationsbaserade modellerna var för grova för att efterlikna Covid-19:s framfart.

Fortsättningvis, den individbaserade nätverksmodellen uppvisar stor potential. För att erhålla en modell med ett högre prediktivt värde, krävs mer arbete med modellkonstruktion samt mer sofistikerade beräkningsresurser än de föregående modellerna för stora populationer. För att runda av, de analytiska modellerna kan under många förhållanden kräva mindre beräkningskraft än nätverksmodellen. Men världen av individbaserad modellering är troligen mer lukrativ i framtiden, då beräkningsresurser troligen inte är lika restriktande.

Referenser

- [1] R. Frerichs, *John snow*, Encyclopedia Britannica, Hämtad: Feb. 01, 2021, 2020. URL: <https://www.britannica.com/biography/John-Snow-British-physician>.
- [2] L. Martcheva, *An Introduction to Mathematical Epidemiology*. New York, NY, USA: Springer, 2015, vol. 61.
- [3] Folkhälsomyndigheten, “Scenarier: Tre smittspridningsscenarier inom regeringsuppdraget *Plan inför eventuella nya utbrott av covid-19*”, Folkhälsomyndigheten, tekn. rapport 20111-1, aug. 2020. URL: <https://www.folkhalsomyndigheten.se/publicerat-material/publikationsarkiv/s/scenarier--tre-smittspridningsscenarier-inom-regeringsuppdraget-plan-infor-eventuella-nya-utbrott-av-covid-19/>.
- [4] J. Panovska-Griffiths, *Coronavirus: We’ve had ‘imperial’, ‘oxford’ and many more models – but none can have all the answers*, april 2020. URL: <https://theconversation.com/coronavirus-weve-had-imperial-oxford-and-many-more-models-but-none-can-have-all-the-answers-135137>.
- [5] R. S. J.M Heffernan och L. Wahl, “Perspectives on the basic reproductive ratio”, *Journal of the Royal Socceity Interface*, årg. 2, s. 281–293, 2005, ISSN: 4. DOI: <http://doi.org/10.1098/rsif.2005.0042>. URL: <https://royalsocietypublishing.org/doi/10.1098/rsif.2005.0042>.
- [6] R. P. Dobrow, *Introduction to Stochastic Processes with R*. John Wiley & Sons, Incorporated, 2016. URL: <https://ebookcentral.proquest.com/lib/chalmers/detail.action?docID=4462510>.
- [7] L. Allen, “A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis”, *Infectious Disease Modelling*, årg. 2(2), s. 128–142, 2017. URL: <https://doi.org/10.1016/j.idm.2017.03.001>.
- [8] N. Van Kampen, “Chapter VIII - The Fokker–Planck Equation”, i *Stochastic Processes in Physics and Chemistry (Third Edition)*, ser. North-Holland Personal Library, Third Edition, Amsterdam: Elsevier, 2007, s. 193–218. DOI: <https://doi.org/10.1016/B978-044452965-7/50011-8>. URL: <https://www.sciencedirect.com/science/article/pii/B9780444529657500118>.
- [9] A. Traulsen, J. C. Claussen och C. Hauert, “Coevolutionary dynamics in large, but finite populations”, *Phys. Rev. E*, årg. 74, s. 011901, 1 juli 2006. DOI: 10.1103/PhysRevE.74.011901. URL: <https://link.aps.org/doi/10.1103/PhysRevE.74.011901>.
- [10] C. M. Macal och M. J. North, “Tutorial on agent-based modeling and simulation”, i *Proceedings of the Winter Simulation Conference, 2005.*, 2005, s. 14–59. DOI: 10.1109/WSC.2005.1574234.
- [11] E. Stattner och N. Vidot, “Social network analysis in epidemiology: Current trends and perspectives”, i *2011 Fifth International Conference on Research Challenges in Information Science*, 2011, s. 1–11. DOI: 10.1109/RCIS.2011.6006866.
- [12] S. Bansal, B. Grenfell och L. A. Meyers, “When individual behaviour matters: Homogeneous and network models in epidemiology”, *Journal of the Royal Society, Interface / the Royal Society*, årg. 4, s. 879–91, nov. 2007. DOI: 10.1098/rsif.2007.1100.
- [13] M. Newman, “The structure and function of complex networks”, *SIAM Review*, årg. 45, nr 2, s. 167–256, 2003. DOI: 10.1137/S003614450342480. URL: <https://doi.org/10.1137/S003614450342480>.
- [14] D. J. Haw, R. Pung, J. M. Read och S. Riley, “Strong spatial embedding of social networks generates nonstandard epidemic dynamics independent of degree distribution and clustering”, *Proceedings of the National Academy of Sciences*, årg. 117, nr 38, s. 23636–23642, 2020. DOI: 10.1073/pnas.1910181117. URL: <https://www.pnas.org/content/117/38/23636>.
- [15] J. A. Firth, J. Hellewell och e. T. Klepac, “Using a real-world network to model localized covid-19 control strategies”, *Nat Med* 26, s. 1616–1622, 2020. DOI: 10.1038/s41591-020-1036-8.

- [16] O. Reich, G. Shalev och T. Kalvari, “Modeling covid-19 on a network: Super-spreaders, testing and containment”, *medRxiv*, 2020. DOI: 10.1101/2020.04.30.20081828. URL: <https://www.medrxiv.org/content/early/2020/05/05/2020.04.30.20081828>.
- [17] A. L. Ziff och R. M. Ziff, “Fractal kinetics of covid-19 pandemic (with update 3/1/20)”, *medRxiv*, 2020. DOI: 10.1101/2020.02.16.20023820. URL: <https://www.medrxiv.org/content/early/2020/03/03/2020.02.16.20023820>.
- [18] B. Blasius, “Power-law distribution in the number of confirmed covid-19 cases”, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, årg. 30, nr 9, s. 93–123, 2020. DOI: 10.1063/5.0013031. URL: <https://doi.org/10.1063/5.0013031>.
- [19] M. Li, J. Chen och Y. Deng, *Scaling features in the spreading of covid-19*, 2020. arXiv: 2002.09199 [physics.soc-ph].
- [20] B. Pettejohn, M. Berryman och M. McDonnell, “Methods for generating complex networks with selected structural properties for simulations: A review and tutorial for neuroscientists”, *Frontiers in Computational Neuroscience*, årg. 5, s. 11, 2011, ISSN: 1662-5188. DOI: 10.3389/fncom.2011.00011. URL: <https://www.frontiersin.org/article/10.3389/fncom.2011.00011>.
- [21] R. Albert och A.-L. Barabási, “Statistical mechanics of complex networks”, *Rev. Mod. Phys.*, årg. 74, nr 1, s. 47–97, jan. 2002. DOI: 10.1103/RevModPhys.74.47. URL: <https://link.aps.org/doi/10.1103/RevModPhys.74.47>.
- [22] T. W. Boonstra, M. E. Larsen och H. Christensen, “Mapping dynamic social networks in real life using participants’ own smartphones”, *Heliyon*, årg. 1, nr 3, e00037, 2015, ISSN: 2405-8440. DOI: <https://doi.org/10.1016/j.heliyon.2015.e00037>. URL: <https://www.sciencedirect.com/science/article/pii/S2405844015300566>.
- [23] L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Skowronski och R. C. Brunham, “Network theory and sars: Predicting outbreak diversity”, *Journal of theoretical biology*, årg. 232, nr 1, s. 71–81, 2005. DOI: 10.1016/j.jtbi.2004.07.026.
- [24] C. M. Hazelbag, J. Dushoff, E. M. Dominic, Z. E. Mthombothi och W. Delva, “Calibration of individual-based models to epidemiological data: A systematic review”, *PLOS Computational Biology*, årg. 16, nr 5, s. 1–17, maj 2020. DOI: 10.1371/journal.pcbi.1007893. URL: <https://doi.org/10.1371/journal.pcbi.1007893>.
- [25] K. Csilléry, M. G. Blum, O. E. Gaggiotti och O. François, “Approximate bayesian computation (abc) in practice”, *Trends in Ecology & Evolution*, årg. 25, nr 7, s. 410–418, 2010, ISSN: 0169-5347. DOI: 10.1016/j.tree.2010.04.001. URL: <https://www.sciencedirect.com/science/article/pii/S0169534710000662>.
- [26] Socialstyrelsen, *Statistik om slutenvård av patienter med covid-19*, Hämtad: 2021-04-29, 2021. URL: <https://www.socialstyrelsen.se/statistik-och-data/statistik/old-statistik-om-covid-19/statistik-om-slutenvard-av-patienter-med-covid-19/>.
- [27] Google, *Rapporter om människors rörlighet*, Hämtad: 2021-04-29, 2021. URL: <https://www.google.com/covid19/mobility/>.
- [28] H. A. Herrmann och J.-M. Schwartz, “Why COVID-19 models should incorporate the network of social interactions”, *Physical Biology*, årg. 17, nr 6, s. 065008, okt. 2020. DOI: 10.1088/1478-3975/aba8ec. URL: <https://doi.org/10.1088/1478-3975/aba8ec>.
- [29] S. Lauer, K. Grantz, Q. Bi, F. Jones, Q. Zheng, H. Meredith, A. Azman, N. Reich och J. Lessler, “The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: Estimation and application”, *Ann Intern Med*, årg. 172, nr 9, s. 577–582, 2020. DOI: 10.7326/M20-0504.
- [30] I. Voinsky, G. Baristaite och D. Gurwitz, “Effects of age and sex on recovery from covid-19: Analysis of 5769 israeli patients”, *The Journal of infection*, årg. 81, nr 2, s. 102–103, 2020. DOI: 10.1016/j.jinf.2020.05.026.

- [31] K. Strålin, E. Wahlström, S. Walther, A. M. Bennet-Bark, M. Heurgren, T. Lindén, J. Holm och H. Hanberger, “Mortality trends among hospitalised covid-19 patients in sweden: A nationwide observational cohort study”, *The Lancet Regional Health - Europe*, årg. 4, s. 100 054, 2021, ISSN: 2666-7762. DOI: <https://doi.org/10.1016/j.lanepe.2021.100054>. URL: <https://www.sciencedirect.com/science/article/pii/S2666776221000314>.
- [32] Folkhälsomyndigheten, “The infection fatality rate of covid-19 in stockholm – technical report”, Folkhälsomyndigheten, tekn. rapport, 2020, Hämtad: 2021-04-23.
- [33] —, “Veckorapport om covid-19, vecka 15”, Folkhälsomyndigheten, tekn. rapport, 2021, Hämtad: 2021-04-29.
- [34] J. Cresson och S. Sonner, “A note on a derivation method for sde models: Applications in biology and viability criteria”, *Stochastic Analysis and Applications*, årg. 36, nr 2, s. 224–239, 2018. DOI: 10.1080/07362994.2017.1386571. eprint: <https://doi.org/10.1080/07362994.2017.1386571>. URL: <https://doi.org/10.1080/07362994.2017.1386571>.
- [35] S. Kang och S. Jung, “Age-related morbidity and mortality among patients with covid-19”, *Infection & chemotherapy*, årg. 52, nr 2, s. 154–164, 2020. DOI: 10.3947/ic.2020.52.2.154.

A Nästa-generationsmetoden

Definiera tillstånden med infekterade individer som element i vektorn $\mathbf{x} = x_i$, $\mathbf{x} \in \mathbb{R}^n$. Infekterade individer är de som har infekterats av en sjukdomsbärare, vilket innebär att exempelvis inkubations tillstånd eller liknande tillhör denna klassifikation. Därefter definiera följande vektorer och matriser

$$\mathcal{F}_i(x_i) := \text{Takten för tillkomsten av nya infekterade till } x_i, \quad (41)$$

$$\mathcal{V}_i(x_i) := (\text{Takten för flödet av individer ut ur } x_i) - (\text{takten för flödet av individer till } x_i), \quad (42)$$

$$\mathcal{F} \in \mathbb{R}^n, \quad \mathcal{V} \in \mathbb{R}^n, \quad (43)$$

$$\frac{d\mathbf{x}}{dt} = \mathcal{F} - \mathcal{V}, \quad (44)$$

$$F := \left. \frac{\partial \mathcal{F}}{\partial \mathbf{x}} \right|_{\mathbf{x}_s}, \quad V := \left. \frac{\partial \mathcal{V}}{\partial \mathbf{x}} \right|_{\mathbf{x}_s}, \quad \mathbf{x}_s\text{-sjukdomsfritt "steady-state" tillstånd, } F, V \in \mathbb{R}^{n \times n}. \quad (45)$$

Från dessa storheter kan generations-matrisoperatoren G , definieras om V är inverterbar som

$$G := FV^{-1}. \quad (46)$$

Utifrån denna matris kan \mathcal{R}_0 definieras som spektralradien för G

$$\mathcal{R}_0 = \max_i \|\lambda\{G\}_i\|, \quad \text{där } \lambda\{G\} := \{\lambda_i \in \mathbb{C} \mid \det(G - \lambda_i I_n) = 0\}. \quad (47)$$

Enligt [2] måste \mathcal{R}_0 uppfylla några krav:

- $\mathcal{R}_0 > 0$, för icke-negativa parametrar,
- $\mathcal{R}_0 = 0$, när sjukdomen ej sprids,
- enheterna måste överensstämna med att \mathcal{R}_0 kan tolkas som antalet sekundära infektioner.

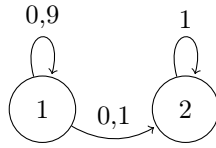
B Markovkedjor

B.1 Övergångsmatrisen \mathcal{P}

Markovkedjor är ett vanligt sätt att simulera slumpvandringar. De Markovkedjor vi kommer att använda kommer ha ett ändligt antal tillstånd. I sådana fall kan man skriva upp en matris som definierar sannolikheten att gå från ett tillstånd till ett annat. Denna matris kallas övergångsmatrisen och betecknas med \mathcal{P} . Elementen i \mathcal{P} kallas övergångssannolikheterna. Nedan ses ett exempel med två tillstånd.

$$\mathcal{P} = \begin{bmatrix} 0,9 & 0,1 \\ 0 & 1 \end{bmatrix}.$$

Notera i detta exempel att radsumman är 1, vilket krävs eftersom de är sannolikheter. Förutom detta se även tillstånd 2 som är ett absorberande tillstånd, ett tillstånd man inte kan lämna. Markovkedjor kan också representeras med hjälp av grafer. I figur 9 ser vi samma information som i \mathcal{P} representerat med en graf.



Figur 9: Exempel på en Markovkedja med två tillstånd, där tillstånd 2 är ett absorberande tillstånd.

B.2 Liten o -notation

$f(x) \in o(g(x))$ innebär att det finns en konstant $k > 0$ så att $0 \leq f(x) < kg(x)$ stämmer för $\forall x \geq a$. Värdet på a får inte bero på värdet på x men får bero på värdet på k . Detta innebär i praktiken att o är en starkare övre gräns än \mathcal{O} . Till exempel så är $\Delta t \notin o(\Delta t)$. Däremot stämmer $\Delta t \in \mathcal{O}(\Delta t)$. Man kan tänka på \mathcal{O} som \leq och o som $<$. Definitionen av o leder till följande gränsvärde vilket behövs i beräkningen av Q

$$\lim_{\Delta t \rightarrow 0^+} \frac{f(\Delta t)}{\Delta t} = 0.$$

B.3 Härledning av Kolmogorovs framåt differentialekvation

Först bildas ett uttryck för sannolikheten för att en övergång skett från tillstånd a till $b = (s, i)$, efter tidssteget Δt :

$$\begin{aligned} \mathcal{P}_{a,(s,i)}(t + \Delta t) &= \underbrace{\mathcal{P}_{a,(s+1,i-1)}(t) \frac{\beta}{N} (s+1)(i-1)\Delta t}_{i \leftarrow (i+1), s \leftarrow (s-1)} + \underbrace{\mathcal{P}_{a,(s,i+1)}(t) \gamma (i+1)\Delta t}_{i \leftarrow (i-1), \Delta s=0} \\ &+ \underbrace{\mathcal{P}_{a,(s,i)}(t) \left[1 - \left(\frac{\beta s i}{N} + \gamma i \right) \Delta t \right]}_{\Delta s = \Delta i = 0} + o(\Delta t). \end{aligned} \quad (48)$$

Sannolikhetstermerna beskriver följande processer:

Infektion:

$\mathcal{P}_{a,(s+1,i-1)}(t)$, beskriver sannolikheten att antalet individer i I -facket har ökat medan antalet i S har minskat. Sett från tidssteget $t + \Delta t$ är antalet individer i facken S, I under tidsteget t : $(S(t) = S(t + \Delta t) + 1, I(t) = I(t + \Delta t) - 1)$.

Tillfrisknande:

$\mathcal{P}_{a,(s,i+1)}(t)$, samma logik gäller för sannolikheten för förflyttningen mellan facken I till R . Nu är antalet individer i facken under tidsteget t mätt från $t + \Delta t$: $(S(t) = S(t + \Delta t), I(t) = I(t + \Delta t) + 1)$.

Ingen ändring:

$\mathcal{P}_{a,(s,i)}(t)$, ingen tillståndsövergång har skett.

Vidare, subtraherar en $\mathcal{P}_{a,(s,i)}(t)$ -term från båda sidor av ekvation (48):

$$\begin{aligned}
\mathcal{P}_{a,(s,i)}(t + \Delta t) - \mathcal{P}_{a,(s,i)}(t) &= \mathcal{P}_{a,(s+1,i-1)}(t) \frac{\beta}{N} (s+1)(i-1)\Delta t + \mathcal{P}_{a,(s,i+1)}(t) \gamma(i+1)\Delta t + \\
&+ \mathcal{P}_{a,(s,i)}(t) \left[1 - \left(\frac{\beta s i}{N} + \gamma i \right) \Delta t \right] - \mathcal{P}_{a,(s,i)}(t) + o(\Delta t) \Leftrightarrow \\
\mathcal{P}_{a,(s,i)}(t + \Delta t) - \mathcal{P}_{a,(s,i)}(t) &= \\
\mathcal{P}_{a,(s+1,i-1)}(t) \frac{\beta}{N} (s+1)(i-1)\Delta t + \mathcal{P}_{a,(s,i+1)}(t) \gamma(i+1)\Delta t + \\
&- \mathcal{P}_{a,(s,i)}(t) \left(\frac{\beta s i}{N} + \gamma i \right) \Delta t + o(\Delta t).
\end{aligned} \tag{49}$$

Vänsterledet i ekvation (49) liknar nu definitionen av derivata för en deriverbar funktion av en variabel:

$$\frac{df}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}. \tag{50}$$

Således genom att multiplicera båda sidorna av likheten i ekvation (49) med $1/\Delta t$ erhålles:

$$\begin{aligned}
\frac{\mathcal{P}_{a,(s,i)}(t + \Delta t) - \mathcal{P}_{a,(s,i)}(t)}{\Delta t} &= \\
\frac{1}{\Delta t} \left[\mathcal{P}_{a,(s+1,i-1)}(t) \frac{\beta}{N} (s+1)(i-1)\Delta t + \mathcal{P}_{a,(s,i+1)}(t) \gamma(i+1)\Delta t \right] + \\
&- \frac{1}{\Delta t} \left[\mathcal{P}_{a,(s,i)}(t) \left(\frac{\beta s i}{N} + \gamma i \right) \Delta t + o(\Delta t) \right].
\end{aligned} \tag{51}$$

Om funktionen $\mathcal{P}_{a,(s,i)}$ är kontinuerlig och om gränsvärdet $\Delta t \rightarrow 0$ existerar medför detta att:

$$\begin{aligned}
\lim_{\Delta t \rightarrow 0} \frac{\mathcal{P}_{a,(s,i)}(t + \Delta t) - \mathcal{P}_{a,(s,i)}(t)}{\Delta t} &= \frac{d\mathcal{P}_{a,(s,i)}}{dt} = \\
\lim_{\Delta t \rightarrow 0} \left[\mathcal{P}_{a,(s+1,i-1)}(t) \frac{\beta}{N} (s+1)(i-1) + \mathcal{P}_{a,(s,i+1)}(t) \gamma(i+1) + \right. \\
&\left. - \mathcal{P}_{a,(s,i)}(t) \left(\frac{\beta s i}{N} + \gamma i \right) + \underbrace{\frac{o(\Delta t)}{\Delta t}}_{\rightarrow 0, \text{ då } \Delta t \rightarrow 0} \right]. \quad \square
\end{aligned} \tag{52}$$

B.4 Generatormatrisen Q

Generatormatrisen Q används för att relatera \mathcal{P} till övergångstakter. Övergångstakterna har till skillnad från övergångsmatrisen med ett beroende på tiden. Q är inte en stokastisk matris utan är istället viktig inom modelleringsaspekten. q_{ij} i Q betecknar takten som ett element rör sig från i till j . Element på diagonalen är definierade som

$$q_{ii} = - \sum_{i \neq j} q_{ij}.$$

Detta leder till att varje rad istället har summan noll. Det betyder att den takt som man tar sig till varje nod måste vara lika stor som takten ut ur noden. Antag nu att vi har en Markovkedja i kontinuerlig tid, $\{X(t)\}_{t \geq 0}$ med $X(0) = i$. Då kommer nästa steg ske vid tid T där $T \sim \lambda_i e^{-\lambda_i t}$. För ett liten tid $\epsilon > 0$, så har vi

$$\mathbb{P}(T_1 < \epsilon) = 1 - e^{-\lambda_i \epsilon} \approx 1 - (1 - \lambda_i \epsilon) = \lambda_i \epsilon$$

där approximationen är en Taylorutveckling. Alltså har vi att för en liten tid Δt så är sannolikheten att lämna tillståndet i ungefär lika med $\lambda_i \Delta t$. Har vi $\mathcal{P}(t)$ kan vi få Q genom att $q_{ij} = \lambda_i p_{ij}$, $\forall i \neq j$ och för q_{ij} med $i = j$ definiera det som ovan. λ är summan av alla övergångstakter för alla möjliga händelser. För \mathcal{P} enligt exemplet ovan får vi att $\lambda_1 = 1$ och $\lambda_2 = 0$.

$$Q = \begin{bmatrix} -0,1\lambda_1 & 0,1\lambda_1 \\ 0\lambda_2 & -0\lambda_2 \end{bmatrix} = \begin{bmatrix} -0,1 & 0,1 \\ 0 & 0 \end{bmatrix}.$$

B.5 Gillespies algoritm

Vi har använt en metod som är vanlig inom biologin, Gillespies algoritm. I Gillespies algoritm vill vi bestämma en stokastisk process med hjälp av två likformigt distribuerade stokastiska variabler, $u_1, u_2 \sim \mathcal{U}[0, 1]$. Praktiskt måste vi först ge initialvärden till modellen. Efter detta kan vi räkna ut sannolikheten för varje händelse. Med dessa sannolikheter och u_1, u_2 kan vi bestämma vilken händelse som skall ske och den andra vid vilken tid. Sedan fortsätter man tills exempel till någon maximal tid eller till ett absorberande tillstånd.

B.6 Härledning av τ

För att kunna simulera τ börjar vi med att notera att T är exponentialfördelad, $T \sim \lambda e^{-\lambda t}$. Definiera nu $F(t) = 1 - e^{-\lambda t} = \mathbb{P}(T \leq t)$, $F(t)$ är alltså den kumulativa distributionen för T . Definiera också \mathcal{U} som den likformiga distributionen på $[0, 1]$. Då får vi

$$F(t) = \mathbb{P}(\mathcal{U} \leq F(t)) = \mathbb{P}(F^{-1}(\mathcal{U}) \leq t)$$

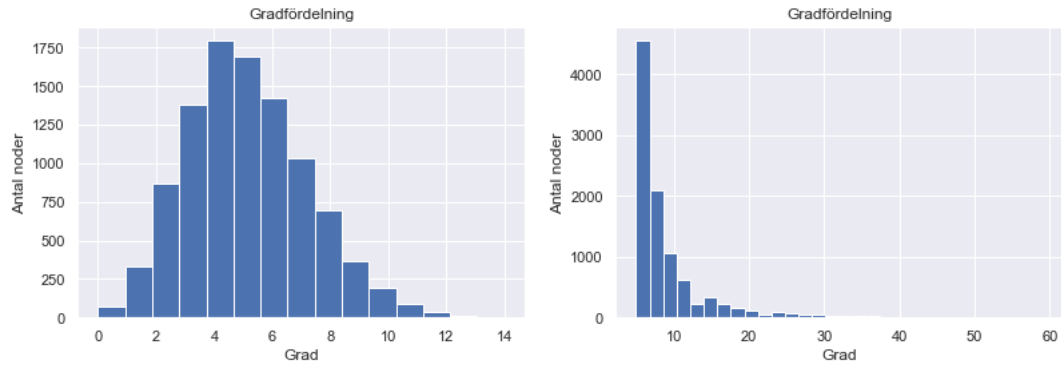
där första likheten kommer från \mathcal{U} s egenskaper och den andra från att $F(t)$ är strikt växande och således inverterbar. Detta innebär att $T = F^{-1}(\mathcal{U})$ och därmed har vi härlett ett uttryck för T . Nu får vi

$$F^{-1}(\mathcal{U}) = -\frac{\ln(1 - \mathcal{U})}{\lambda} = -\frac{\ln(\mathcal{U})}{\lambda}$$

där den första likheten är uträkning av inversen och andra från det faktum att \mathcal{U} och $1 - \mathcal{U}$ har samma distribution.

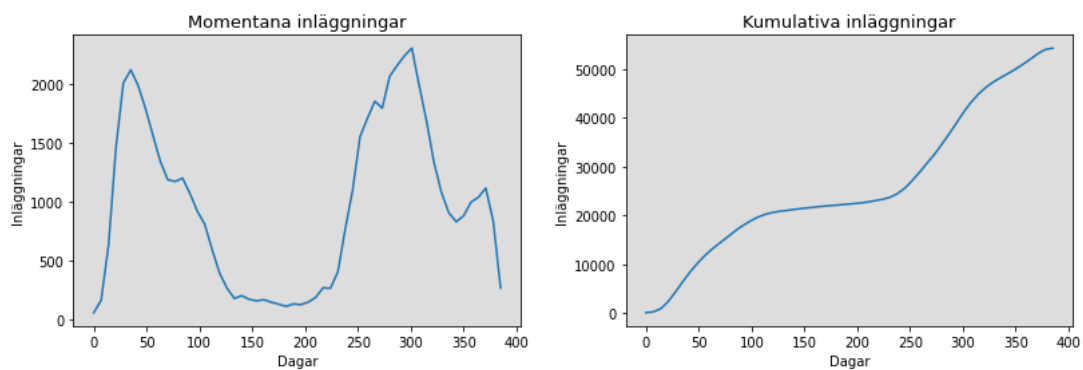
C Figur gradfördelning

Figur 10 presenterar hur fördelningen av grader skiljer sig mellan en Erdős-Renyi-graf och en Barabási-Albert-graf.

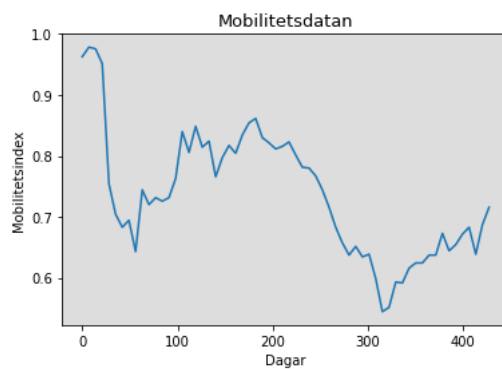


Figur 10: Fördelning av grader till en Erdős-Renyi graf och till en Barabási-Albert graf med 10000 noder.

D Data



Figur 11: Linjärt interpolerad inläggningsdata som en funktion av dagar, given som både momentant och kumulativt antal inläggningar.



Figur 12: Linjärt interpolerad mobilhetsdata som en funktion av dagar.

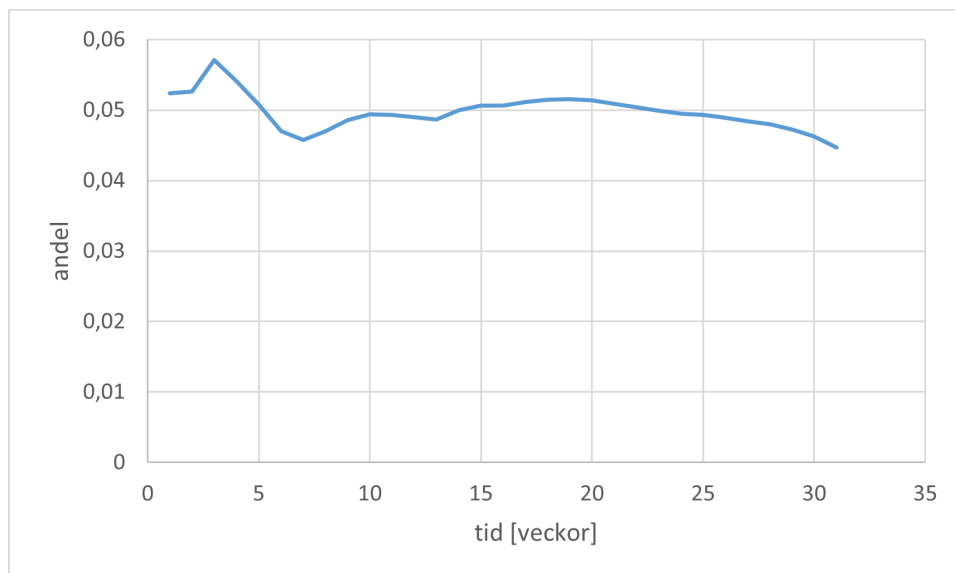
E Länk till koden

Fullständig kod finns på GitHub:

<https://github.com/ElinaPettersson/MVEX01-21-25-kandidatarbete.git>

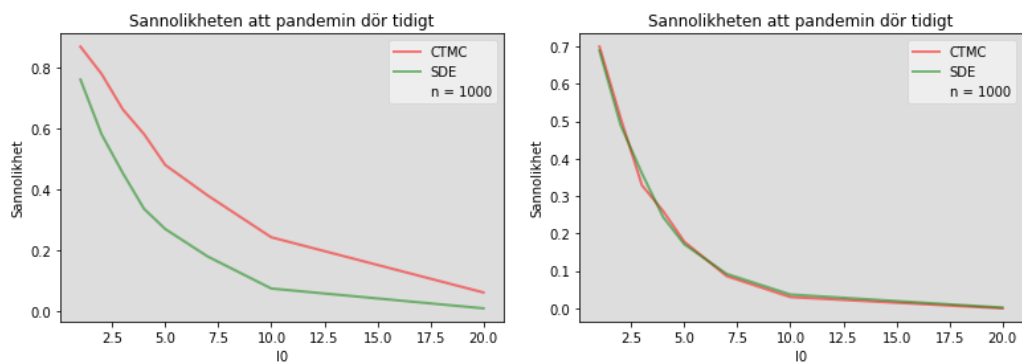
F Uppskattning av *hospitalPar*

I nätverksmodellen avgörs sjukhusinläggningar av parametern *hospitalPar*, vilken uppskattades från litteratordata från Socialstyrelsen och Folkhälsomyndigheten. Data mellan oktober 2020 och april 2021 summerades veckovis och andel mellan slutenvårdade och bekräftade fall beräknades. Andelen låg runt 5 % med låg varians, se figur 13.



Figur 13: Andelen mellan kumulativa sjukhusinläggningar och kumulativa bekräftade sjuka.

G Sannolikheten att pandemin dör tidigt



(a) Graf över proportionen av de stokastiska modellerna som dog tidigt som en funktion av I_0 , lerna som dog tidigt som en funktion av I_0 , här är parametrarna $N_{Markov} = 10^4$, $\beta_{Markov} = 0,075$, $N_{SDE} = 10^7$, $\beta_{SDE} = 0,09$, $\gamma = 1/14$. (b) Graf över proportionen av de stokastiska modellerna som dog tidigt som en funktion av I_0 , lerna som dog tidigt som en funktion av I_0 , här är parametrarna $N = 10^4$, $\beta = 0,1$, $\gamma = 1/14$ för båda modellerna.

Figur 14: Grafer över sannolikheten att pandemin dör tidigt

H Spridningen för Markovprocess modellen plottat för några realiseringar

Population	σ_{min}	$\bar{\sigma}$	σ_{max}
$N = 10^3$	1,6171	3,9274	10,6911
$N = 10^4$	2,8635	7,8247	33,4471
$N = 10^5$	5,1713	12,4773	91,1430

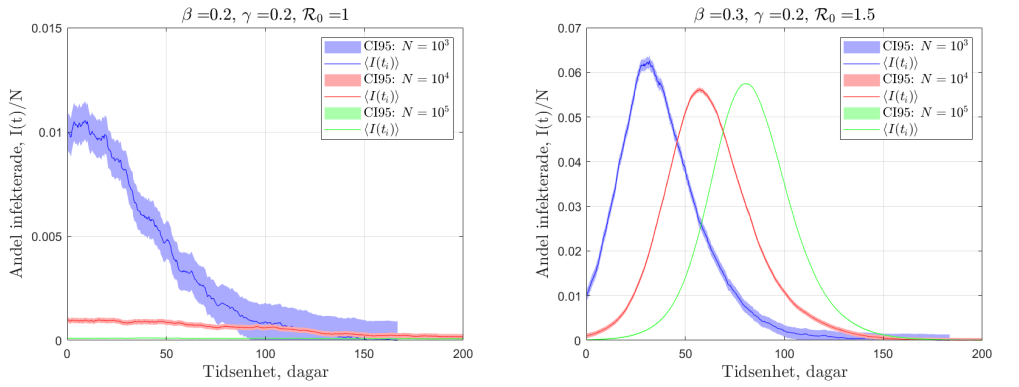
(a) $\mathcal{R}_0 = 1$

Population	σ_{min}	$\bar{\sigma}$	σ_{max}
$N = 10^3$	3.2012	11.1313	25,2591
$N = 10^4$	19.6391	71,4215	302,8599
$N = 10^5$	86.8145	535,3230	2246,2729

(b) $\mathcal{R}_0 = 1,5$

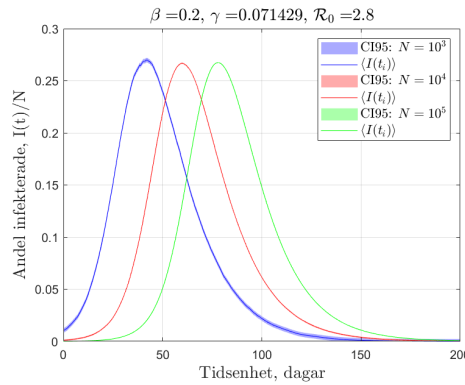
Population	σ_{min}	$\bar{\sigma}$	σ_{max}
$N = 10^3$	3,8694	16,2818	48.2553
$N = 10^4$	21.1434	130,2228	456,0727
$N = 10^5$	190,5288	1268,9684	5321,0357

Tabell 2: $\mathcal{R}_0 = 2,8$



(a) \mathcal{R}_0 som svarar mot 1 sekundär infektion.

(b) \mathcal{R}_0 som svarar mot 1,5 sekundära infektioner.



(c) \mathcal{R}_0 som svarar mot 2,8 sekundära infektioner.

Figur 15: Spridningsfigurer för Markovmodellen.

Samtliga simulationer utfördes 100 gånger, $\bar{\sigma}$ baseras därför på medelvärdet för samtliga simuleringar, σ är den skattade standardavvikelsen från medelvärdet $\langle I \rangle$.

I Spridningen för Itô-SDE modellen plottat för några realiseringar

Population	σ_{min}	$\bar{\sigma}$	σ_{max}
$N = 10^3$	1,3624	3,7058	8,9488
$N = 10^4$	4,0037	9,7498	46,5847
$N = 10^5$	5,0181	12,3962	88,9573

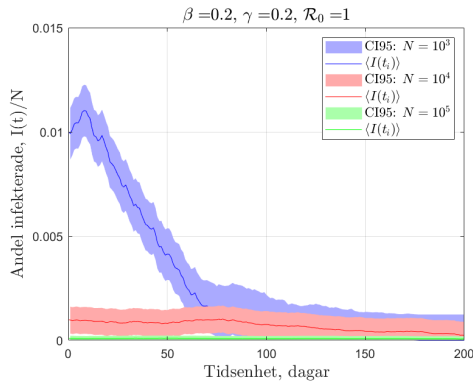
(a) $\mathcal{R}_0 = 1$

Population	σ_{min}	$\bar{\sigma}$	σ_{max}
$N = 10^3$	3,7268	10,7151	29,4023
$N = 10^4$	15,8532	79,0221	249,7911
$N = 10^5$	154,0435	687,2719	2508,8707

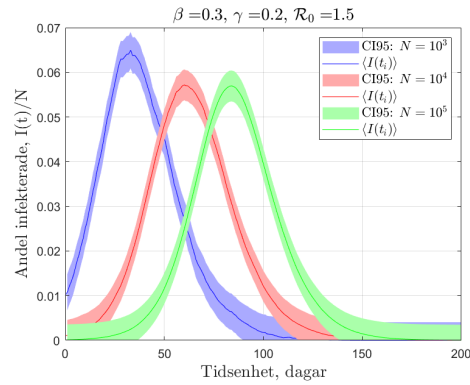
(b) $\mathcal{R}_0 = 1,5$

Population	σ_{min}	$\bar{\sigma}$	σ_{max}
$N = 10^3$	5,5467	17,9232	48,7982
$N = 10^4$	23,1229	138,0374	491,0957
$N = 10^5$	181,8157	1359,8455	4292,0552

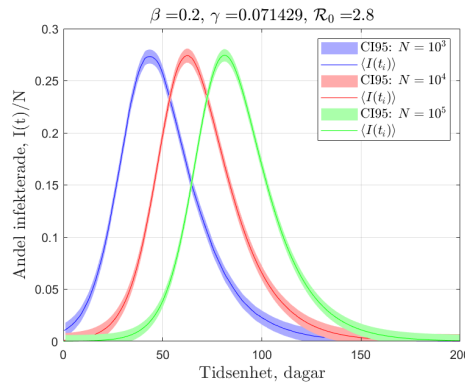
Tabell 4: $\mathcal{R}_0 = 2,8$



(a) \mathcal{R}_0 som svarar mot 1 sekundär infektion.



(b) \mathcal{R}_0 som svarar mot 1,5 sekundära infektioner.



(c) \mathcal{R}_0 som svarar mot 2,8 sekundära infektioner.

Figur 16: Spridningsfigurer för SDE:n.

Samtliga simulationer utfördes 100 gånger, $\bar{\sigma}$ baseras därför på medelvärdet för samtliga simuleringar, σ är den skattade standardavvikelsen från medelvärdet $\langle I \rangle$.