



Stokastiska modeller för evolution

En studie av fixeringsannolikhet i cancermodeller

Stochastic models for evolution

Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet

Kandidatarbete inom civilingenjörsutbildningen vid Chalmers tekniska högskola

Ida Berg

Elin Edgren

Jonathan Kedner

Sofia Larsson

Johanna Lundström

Matilda Selvaraj Tivesten

Stokastiska modeller för evolution

En studie av fixeringsannolikhet i cancermodeller

Examensarbete för kandidatexamen i matematisk statistik vid Göteborgs universitet
Elin Edgren Jonathan Kedner

Kandidatarbete i matematik inom civilingenjörsprogrammet Bioteknik vid Chalmers
Ida Berg Sofia Larsson Johanna Lundström

Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers
Matilda Selvaraj Tivesten

Handledare: Philip Gerlee

Institutionen för Matematiska vetenskaper
CHALMERS TEKNISKA HÖGSKOLA
GÖTEBORGS UNIVERSITET
Göteborg, Sverige 2023

Förord

Detta kandidatarbete är skrivet vid Institutionen för Matematiska vetenskaper på Chalmers tekniska högskola och Göteborgs universitet och handlar om att modellera mikroevolution stokastiskt. Vi vill framför allt tacka vår handledare Philip Gerlee som hjälpt oss mycket i arbetet. Han har bland annat hållit i föreläsningar som underlättat för att förstå ämnet och litteraturen vi läst, och även hjälpt till med att lösa problem som har uppstått under projektets gång. Dessutom har han gett oss feedback under tiden och hjälpt oss att granska rapporten.

Vi har till stor del arbetat tillsammans för att ta fram information och skriva rapporten, och vi har gemensamt skapat modellerna. I tabellen nedan listas de främsta författarna för varje del i rapporten.

Bidragsrapport

§	Rubrik	Författare
	Förord	Ida, Sofia
	Populärvetenskaplig presentation	Matilda, Jonathan
	Sammandrag	Ida
	Abstract	Ida
1	Inledning	Ida, Sofia
1.1	Syfte	Sofia, Ida
1.2	Problemformulering	Sofia
1.3	Avgränsningar	Sofia
1.4	Samhälleliga och etiska aspekter	Elin, Ida
2	Teori	
2.1	Selektion	Ida, Elin
2.1.1	Frekvensberoende selektion	Ida, Johanna
2.2	Moranprocess	Johanna, Matilda
2.2.1	Moranprocess utan selektion	Johanna, Matilda
2.2.2	Moranprocess med konstant selektion	Johanna, Matilda
2.2.3	Moranprocess med frekvensberoende selektion	Johanna, Matilda
2.3	Selektion hos cancerceller	Ida, Sofia
2.3.1	Modell för cancertillväxt	Ida, Sofia
2.4	Absorptionssannolikhet	Sofia
2.4.1	Absorptionssannolikhet utan selektion	Sofia
2.4.2	Absorptionssannolikhet med konstant selektion	Johanna, Sofia
2.4.3	Absorptionssannolikhet med frekvensberoende selektion	Ida, Johanna
3	Metod	
3.1	Modellering av Moranprocessen	Sofia, Matilda
3.1.1	Modell 1 - Moranprocess utan selektion	Sofia
3.1.2	Modell 2 - Moranprocess med konstant selektion	Sofia
3.1.3	Modell 3 - Moranprocess med frekvensberoende selektion	Sofia
3.2	Modell 4 - Modell för cancertillväxt	Ida, Johanna
4	Resultat	
4.1	Modell 1 - Moranprocess utan selektion	Elin
4.2	Modell 2 - Moranprocess med konstant selektion	Sofia
4.3	Modell 3 - Moranprocess med frekvensberoende selektion	Johanna
4.4	Modell 4 - Modell för cancertillväxt	Ida, Johanna
5	Diskussion	Ida
5.1	Diskussion av resultat	Elin, Ida
5.2	Diskussion av metodval och avgränsningar	Sofia, Jonathan
5.3	Slutsats	Sofia, Elin
A	Appendix - Notationslista	Ida, Johanna
B	Appendix - Definitioner och härledningar	
B.1	Markovkedjor	Johanna, Jonathan
B.2	Konfidensintervall	Jonathan, Elin
B.3	Härledning av generell formel för absorptionssannolikhet	Johanna
C	Appendix - Källkod	
C.1	Modell 1 - Moranprocess utan selektion	Matilda
C.2	Modell 2 - Moranprocess med konstant selektion	Matilda
C.3	Modell 3 - Moranprocess med frekvensberoende selektion	Matilda
C.4	Modell 4 - Modell för cancertillväxt	Sofia
D	Appendix - Rådata	
D.1	Modell 1 - Moranprocess utan selektion	Ida
D.2	Modell 2 - Moranprocess med konstant selektion	Ida
D.3	Modell 3 - Moranprocess med frekvensberoende selektion	Ida
D.4	Modell 4 - Modell för cancertillväxt	Ida, Johanna

Övriga bidrag

Utöver rapportskrivande har projektarbetet även inneburit en del praktiskt arbete som inte visas i bidragsrapporten. Nedan beskrivs därför alla övriga bidrag till projektet.

Planering

Planering och upplägg av projektet gjordes till största del i samspel med handledaren. Gruppen har haft regelbundna möten varje vecka, både inom gruppen och med handledaren. Under gruppmöten diskuterades vad projektet skulle innebära samt hur arbetet skulle läggas upp och fördelas inom gruppen. Under mötena har även ett mötesprotokoll förts för att lättare hålla reda på beslut som tagits och generellt kunna reflektera över hur långt man kommit i projektet. Planeringsrapporten diskuterades och skrevs gemensamt och skapandet av tidsplan bidrog alla i gruppen till.

Under arbetets gång har det förts en loggbok för varje persons insats. Loggboken består av en tidslogg och en gemensam dagbok. Den tid man har lagt ner varje dag har loggats i tidsloggen tillsammans med en kort förklaring. I dagboken har en längre förklaring till arbetet man gjort skrivits varje vecka.

Genomförande

Arbetet inleddes med att Matilda påbörjade kod för simulering av Moranprocessen. Jonathan och Elin skrev kod för beräkning av konfidensintervall och genererandet av grafer som sedan Sofia omarbetade. Sofia implementerade även de analytiska lösningarna. Vid upptäckt av att stegsimuleringen i Markovkedjan blivit fel, där de analytiska och simulerade lösningarna för Modell 2 inte överensstämde, lade alla i gruppen ned tid på att felsöka problemet. Johanna och Matilda lyckades lösa det. För Modell 3 uppstod ett problem med den approximerade analytiska lösningen. Elin och Jonathan löste problemet genom att använda en numerisk lösning istället. Elin simulerade och satte parametervärden för att illustrera selektionens påverkan på resultatet för Modell 1 och 2. Elin, Matilda, Ida och Johanna studerade valet av parametervärden för Modell 3 och undersökte jämviktspunkter samt tredjedelsregeln. Matilda omarbetade därefter koden och renskrev den för Modell 1, 2 och 3.

Då tredjedelsregeln inte fungerar för det valda spelet i rapporten undersökte Elin ännu ett spel, som Matilda skrev kod för. Detta för att försöka förstå vad den skulle innebära. Dock togs inte denna modell med i rapporten då det blev påtänkt lite sent i projektet och inte ansågs vara relevant för syftet.

Ida, Johanna och Sofia skrev koden och valde parametervärdena för Modell 4 där lite av koden återanvändes från Moranprocessmodellerna. Johanna skrev kod för färgdiagram som Sofia sedan implementerade. Vid upptäckt att koden gick väldigt långsamt lade Ida, Johanna och Sofia ned tid på att försöka få den att gå snabbare. Därefter omarbetade Sofia koden och renskrev den.

Utöver de källor som vi har fått från vår handledare har Ida sökt efter källorna i inledningen samt för avsnitten om cancer. Elin sökte efter källor i avsnittet samhälleliga och etiska aspekter som sedan inte användes. Johanna har härlett det generella uttrycket för absorptionssannolikhet som återfinns i Avsnitt B.3 och Elin har räknat på härledning av de analytiska lösningarna av absorptionssannolikheten för Modell 1 och 2 där härledningen för Modell 1 sedan inte inkluderades i rapporten på grund av överflöd. Jonathan har skrivit matematiska definitioner som sedan beslutades att tas bort från rapporten på grund av överflöd. Matilda har ritat bilderna som finns med i teoriavsnittet. Sofia och Johanna undersökte och fixade problem med referenslistan. För metodavsnittet diskuterades det mycket inom gruppen om hur den skulle skrivas. Därmed lade Elin, Jonathan och Sofia ned mycket tid på att skriva olika varianter av metodavsnittet. Alla har hjälps åt att generera nytt resultat i form av figurer och tabeller under arbetets gång när koden uppdaterats eller bytt utseende på graferna.

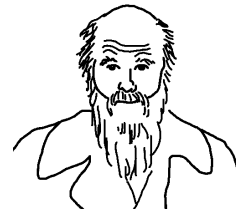
Utöver de specifika författaravsnitten i bidragsrapporten har hela gruppen gemensamt diskuterat rapportens innehåll under arbetets gång. Revidering av rapporten har främst gjorts av Sofia, Ida, Johanna och Elin.

Diskussionsbidrag

Inför skrivandet av diskussionen diskuterade alla i gruppen under gemensamma möten vad som eventuellt skulle kunna skrivas baserat på resultatet, metoden och teorin. Resultatet diskuterades även med handledaren för att få fler idéer till diskussionen.

Populärvetenskaplig presentation

Många känner säkert till namnet Darwin. Han var en pionjär inom evolution och hans teorier kan sammanfattas med uttrycket "survival of the fittest", det vill säga att den bäst anpassade överlever. Detta uttryck fungerar väl när man tittar på stora populationer. När man tittar på en liten population av individer så behöver det dock inte nödvändigtvis vara så att den bäst anpassade överlever. I vårt kandidatarbete har vi studerat detta fenomen genom att simulera cancertillväxt enligt fyra olika modeller.



Charles Darwin

I grund och botten är cancertillväxt en slumpmässig process. Dels sker mutationerna, som skapar cancerceller, slumpmässigt och dels styr slumpen om cancercellerna delar sig och ökar i antal. Cancer kan därför modelleras som en slumpmässig process, som inom matematiken kallas för en stokastisk process. Vanligtvis använder man sig av deterministiska modeller för att modellera evolution. Determinism innebär att resultatet inte styrs av slumpen alls, och är alltså motsatsen till stokastik. I en deterministisk modell är det den som är bäst anpassad som alltid tar över populationen. Detta är fallet i en stor population eftersom små slumpmässiga förändringar har liten inverkan då och oftast mer eller mindre tar ut varandra. Därför lämpar sig deterministiska modeller väl för att beskriva stora populationer. I en liten population däremot, har slumpen en mycket mer betydande inverkan. Då gäller inte alltid "survival of the fittest". Därför är det särskilt relevant att använda stokastiska modeller för att beskriva cancertillväxt då det är en relativt liten population av celler som studeras. De stokastiska modellerna tar till exempel hänsyn till att den första cancercellen hade kunnat dö innan den hinner dela sig, även i de fall där den har ett stort övertag.

Med våra modeller simulerar vi hur cellerna i en population slumpmässigt delar på sig och dör. Vi har studerat fixerings sannolikheten för fyra olika modeller av stigande komplexitet under olika antaganden om cellernas egenskaper. Fixerings sannolikheten är sannolikheten att en ensam cancercell i en population av friska celler tar över hela populationen, alltså att hela populationen tillslut endast består av cancerceller. Genom att upprepade gånger simulera vad som händer då vi introducerar en cancercell i en population av friska celler kan vi uppskatta fixerings sannolikheten. I våra modeller bortser vi från slumpmässiga mutationer och cellerna förökar sig genom delning. Detta innebär att det måste finnas minst en cell av en viss typ för att det ska kunna bli fler av den typen.

I de tre första modellerna utgår vi från att det hela tiden finns lika många celler i populationen. Om en cell delar på sig så måste därmed en annan cell dö samtidigt för att antalet inte ska ändras. I den första modellen räknar vi med att cancercellerna har samma förutsättningar för att dela sig och samma risk att dö som de vanliga cellerna. Då har alla typer av celler lika stor chans att ta över hela populationen. I den andra modellen har cancercellerna istället en högre reproduktionstakt än de friska cellerna, vilket innebär att de har bättre förutsättningar att dela sig. Då får de ett övertag som gör att fixerings sannolikheten är högre i den andra modellen än i den första. Fixerings sannolikheten blir större ju högre reproduktionstakten är.

I den tredje modellen beror cellernas reproduktionstakt på hur stor andel av populationen som är cancerceller. Beroende på cancercellernas egenskaper kan de få en fördel som beror på antalet cancerceller som redan finns i populationen. Cancercellerna kan exempelvis ha en egenskap som gynnar reproduktionstakten i en population då det finns få cancerceller. Om cancercellerna blir för många påverkar inte egenskapen reproduktionstakten längre. Detta kallas för frekvensberoende selektion eftersom selektionen beror på frekvensen, alltså andelen, av cancerceller i populationen. Beroende på fördelen som cancercellerna får kan fixerings sannolikheten antingen vara högre eller lägre än fallet när de inte konkurrerar med de friska cellerna.

I den fjärde och sista modellen modellerar vi bröstcancer. Om man tittar på ett tvärsnitt av en frisk bröstkanal så ligger cellerna i en ring med ett hålrum i mitten. Cancerceller bryr sig inte om att hålla sig till den ordnade ringen, utan kan även växa in i hålrummet. Tillslut det blir för fullt innanför ringen och då kan bröstkanalen spricka, vilket låter cancer sprida sig till resten av kroppen.

I den sista modellen finns det tre stora skillnader mot de tidigare modellerna. För det första räknar vi inte längre med att populationsstorleken är konstant. Det totala antalet celler kan alltså variera över tiden. För det andra tar vi även hänsyn till var i ringen som cellerna är i relation till varandra. De friska cellerna måste vara en del av ringen, och om en frisk cell delar på sig utan att det finns plats för den nya cellen så dör den nya cellen. Till sist beaktar vi även en bärformåga hos cellerna. Bärformågan är det maximala antalet celler som kan samexistera på de begränsade resurserna som finns, till exempel plats och näring. Man kan likna det vid antalet säten på en buss, där de friska cellerna är passagerare som är väldigt noga med att ha säkerhetsbälte på sig. Om det kommer en ny frisk cell och den inser att det inte finns en sittplats ledig så går den genast av bussen.

En egenskap som cancerceller har är att kunna öka bärformågan. Ibland ökas även bärformågan för angränsande celler i någon utsträckning. Man kan se det som att cancercellerna är passagerare som inte har samma pliktrogenhet till säkerhetsbälten som de friska cellerna har, och de bryr sig inte heller om sociala normer. Om en cancercell går på bussen och inser att det inte finns något ledigt säte, sätter den sig istället i närmaste främlings knä. På så sätt kan cancercellerna växa på varandra i bröstkanalen och täppa igen den.

I de första tre modellerna har vi studerat reproduktionstaktens påverkan på fixeringssannolikheten, medan i den sista modellen har vi även studerat påverkan av bärformågan. Den stora frågan är vilken av dessa faktorer som har störst påverkan. Vi simulerade hundratals förlopp av cancertillväxt för olika värden på reproduktionstakt och bärformåga för att kunna jämföra påverkan. Vi såg då att bärformågan har stor påverkan på cancercellerna. Trots att cancercellerna delar med sig av sin bärformåga höjs fixeringssannolikheten med ökad bärformågan. Redan då bärformågan börjar behållas lite grann påverkas fixeringssannolikheten mycket. Detta tyder på att det räcker att cancercellerna endast behåller en liten del av bärformågan till sig själva för att de ska få en stor fördel av det.

Det vi har kommit fram till är att under de antaganden som vi gjort i våra modeller så är bärformågan och dess fördelning på angränsande celler en stor faktor som påverkar hur aggressiv en tumör är. Slutligen är dessa modeller bara modeller. De kan ge oss en förståelse för hur olika egenskaper kan påverka tumörutveckling, särskilt i ett tidigt skede, men de kan inte ta hänsyn till alla faktorer som kan komma att påverka utvecklingen av sjukdomen.

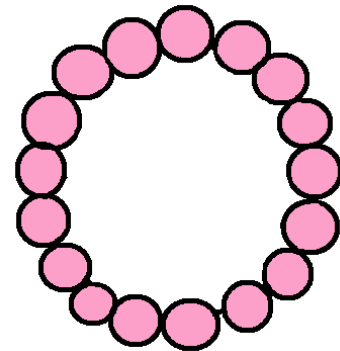


Illustration av celler i tvärsnittet av en frisk bröstkanal.

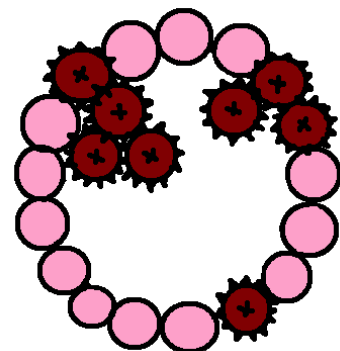


Illustration av hur bröstkanalen invaderas av cancerceller.

Sammandrag

Cancer är en välkänd och förödande sjukdom som drabbar många personer varje år. Denna sjukdom är svårbehandlad då cancerbehandling främjar evolution av resistens hos cancercellerna, vilket leder till allt mer svårbehandlad cancer. Att studera och förstå cancercellers evolutionsförlopp är därför en viktig forskning som berör många. Syftet med detta projekt är att skapa stokastiska modeller som undersöker påverkan av modellparametrarna reproduktionstakt, bärformåga och populationsstorlek för att kunna studera hur evolutionsförlopp ter sig på olika nivåer av komplexitet. Resultatet tas fram som sannolikheten att den ena genotypen, som representerar en cancercell, fixeras och alltså tar över populationen.

För att göra detta har de första modellerna baserats på en Moranprocess. De olika modellerna har studerat olika varianter av selektion vilka är neutral evolution utan selektion, konstant selektion och frekvensberoende selektion, alla med en konstant population. Resultatet från dessa modeller visade att en större reproduktionstakt bidrar till en större fixeringssannolikhet i populationen jämfört med ett neutralt fall. Utöver dessa modeller så skapades även en modell med varierande populationsstorlek och selektion för att beskriva cancertillväxt, där selektionen utöver reproduktionstakt dessutom tagit hänsyn till cellens bärformåga. Resultatet från cancermodellen visade att både reproduktionstakt och bärformåga spelar in vid fixering av cancerceller i en population. Reproduktionstakten vägde lite tyngre när bärformågan delades med omgivningen medan om bärformågan behölls till den egna cellen så vägde den så tungt att reproduktionstakten inte spelade någon större roll. Alla modeller har flera begränsningar som gör att resultaten inte är helt tillförlitliga att applicera på verkligheten. De kan dock ändå skapa en förståelse för hur tumörförlopp ter sig och ge insikt och inspiration för vidare forskning kring det.

Abstract

Cancer is a well-known and devastating disease that affects many individuals every year. This disease is difficult to treat because cancer treatment promotes evolution of resistance in cancer cells, which leads to increasingly intractable cancer. Studying and understanding the evolutionary process of cancer cells is therefore important research that has an impact on many people. The purpose of this project is to create stochastic models that investigate the influence of the model parameters reproduction rate, carrying capacity and population size to study how evolutionary processes appear at different levels of complexity. The results are expressed as the probability that one genotype, who represents a cancer cell, is fixated and thus takes over the population.

This was done by basing the first models on a Moran process. The different models have described different variants of selection which are no selection, constant selection, and frequency dependent selection, all with a constant population. The results from these models showed that a greater reproduction rate contributes to a larger fixation probability in the population compared with a neutral case. In addition to these models, a more complex model of tumour growth with varying population size and selection was also created, where the selection besides reproduction rate also took carrying capacity into account. The results from the cancer model growth showed that both reproduction rate and carrying capacity has an influence on the fixation probability of cancer cells in a population. The reproduction rate had a larger impact when the carrying capacity was shared with the surrounding, while if it was kept to the cell the carrying capacity had such a large impact that the reproduction rate did not matter. All models have several limitations which makes the results not completely trustworthy to apply in reality. However, they can still create an understanding of tumour evolution and provide insight and inspiration for further research.

Innehåll

1	Inledning	1
1.1	Syfte	1
1.2	Problemformulering	2
1.3	Avgränsningar	2
1.4	Samhälleliga och etiska aspekter	2
2	Teori	4
2.1	Selektion	4
2.1.1	Frekvensberoende selektion	4
2.2	Moranprocess	5
2.2.1	Moranprocess utan selektion	6
2.2.2	Moranprocess med konstant selektion	7
2.2.3	Moranprocess med frekvensberoende selektion	7
2.3	Selektion hos cancerceller	7
2.3.1	Modell för cancertillväxt	8
2.4	Absorptionssannolikhet	9
2.4.1	Absorptionssannolikhet utan selektion	9
2.4.2	Absorptionssannolikhet med konstant selektion	9
2.4.3	Absorptionssannolikhet med frekvensberoende selektion	10
3	Metod	11
3.1	Modellering av Moranprocessen	11
3.1.1	Modell 1 - Moranprocess utan selektion	11
3.1.2	Modell 2 - Moranprocess med konstant selektion	11
3.1.3	Modell 3 - Moranprocess med frekvensberoende selektion	11
3.2	Modell 4 - Modell för cancertillväxt	12
4	Resultat	13
4.1	Modell 1 - Moranprocess utan selektion	13
4.2	Modell 2 - Moranprocess med konstant selektion	13
4.3	Modell 3 - Moranprocess med frekvensberoende selektion	14
4.4	Modell 4 - Modell för cancertillväxt	15
5	Diskussion	18
5.1	Diskussion av resultat	18
5.2	Diskussion av metodval och avgränsningar	20
5.3	Slutsats	20
A	Appendix - Notationslista	i
B	Appendix - Definitioner och härledningar	ii
B.1	Markovkedjor	ii
B.2	Konfidensintervall	ii
B.3	Härledning av generell formel för absorptionssannolikhet	ii
C	Appendix - Källkod	iv
C.1	Modell 1 - Moranprocess utan selektion	iv
C.2	Modell 2 - Moranprocess med konstant selektion	viii
C.3	Modell 3 - Moranprocess med frekvensberoende selektion	xii
C.4	Modell 4 - Modell för cancertillväxt	xviii

D	Appendix - Rådata	xxvi
D.1	Modell 1 - Moranprocess utan selektion	xxvi
D.2	Modell 2 - Moranprocess med konstant selektion	xxvi
D.3	Modell 3 - Moranprocess med frekvensberoende selektion	xxvi
D.4	Modell 4 - Modell för cancertillväxt	xxvii

1 Inledning

En vanlig uppfattning om evolution är att det betyder "survival of the fittest" (jfr sv. den starkes överlevnad), där den individ som är bäst anpassad är den som kommer att överleva, vilket är en förenkling av verkligheten. Evolution innebär att en population förändras och utvecklas över tid, där många faktorer påverkar processen såsom populationsstorlek och tidsaspekter. Genom att matematiskt analysera evolution kan förståelsen för de olika evolutionära processer som sker i universum utökas. Processerna kan vara allt från långsiktig evolution för att förstå utvecklingen från det förflutna, till att studera kortsiktig evolution av exempelvis sjukdomsförlopp, såsom cancer.

Begreppen mikroevolution och makroevolution är två undergrupper inom evolution. Det är omdiskuterat vad begreppen egentligen innebär. Huvudsakligen sker makroevolution över en längre tid med stora evolutionära förändringar mellan olika arter, medan mikroevolution sker under en kortare tid med evolutionära förändringar i mindre skala inom arten [1]. Mikroevolution kan dessutom ske i vilken riktning som helst, till skillnad från makroevolution som generellt går i den riktning som är optimal för organismens överlevnad och fortplantning [1]. Detta gör även att förändringarna för mikroevolution är reversibla, medan makroevolution är en irreversibel process [2]. Då mikroevolution sker inom arter påverkas den också av fenomen som finns inom populationer såsom mutationer, selektion, genetisk drift och genflöde [3]. Detta projekt kommer fokusera på mikroevolution och endast beakta selektion och genetisk drift.

En förutsättning för evolution är som sagt att det kan uppstå mutationer i celler då mutationerna kan leda till förbättrade egenskaper i genupsättningen. Ett exempel på konsekvens av denna process är cancer. I cancer har muterade celler bland annat en hög reproduktionstakt och kringgår därför ett flertal av kroppens säkerhetsfunktioner som ska begränsa ohämmad celledelning [4]. På så sätt kan cancerceller undgå att genomgå programmerad celledöd när de inte betar sig som de ska och således växa till att bilda en tumör. Eftersom cancerceller inte självdör som de borde så ser deras evolutionära process annorlunda ut jämfört med andra evolutionära processer [4]. Generna inaktiveras och modifieras utan att cancercellens reproduktionstakt försämras, vilket gör att cancercellen kan förbättra sina egenskaper på kort tid. Cancertillväxt kan därför ses som en destruktiv evolution som medför att de funktioner som är till för att skydda organismer mot cancer inte längre fungerar [4].

För att modellera och analysera evolution måste hänsyn tas till vilka aspekter som studeras. Ofta är en deterministisk modell lämplig för att studera evolution över lång tid [4]. Dock, som tidigare nämnts, är mikroevolution en slumpmässig process eftersom den kan ske i vilken riktning som helst. En effekt av detta är den genetiska driften, där förändringarna som sker inte påverkas av vilka förändringar som är av fördel eller nackdel för individen. För en biologisk mikroevolution lämpar det sig därför bättre att analysera processen med stokastiska modeller. I detta projekt studeras därför utbredning av en genotyp i en population genom fyra stokastiska modeller med ökande komplexitet. De tre första modellerna beaktar en konstant populationsstorlek, vilket innebär att avdöningstakten och reproduktionstakten är lika och därmed ger ett konstant antal individer i populationen. Den fjärde modellen studerar en varierande populationsstorlek, vilket innebär att populationsstorleken kan variera beroende på förhållandet mellan avdöningstakten och reproduktionstakten. Selektion varierar också mellan modellerna där de enklaste fallen är helt utan selektion samt med en konstant selektion. Selektionen kan sedan kompliceras genom att modellera en frekvensberoende selektion och slutligen att den varierar beroende på reproduktionstakt och bärformåga. Den slutliga modellen ska sedan kunna användas för att simulera tumörförlopp i kroppen och på så sätt samla kunskap som kan tillämpas inom cancerforskning.

1.1 Syfte

Syftet med projektet är att studera evolutionsförlopp på mikroskala för att undersöka hur de olika modellparametrarna populationsstorlek, reproduktionstakt och bärformåga påverkar den evolutionära processen. Uppgiften är att ta fram stokastiska modeller, med en ökande komplexitet, för att till slut skapa en modell som efterliknar förloppet av hur cancerceller tar över en population av vanliga celler. Modellerna utvärderas och jämförs genom att studera absorptionssannolikheten,

som visar hur stor sannolikheten är att en av genvarianterna tar över hela populationen, och hur den beror på modellparametrarna.

1.2 Problemformulering

Projektets uppgift är att undersöka hur modellparametrarna i de framtagna stokastiska modellerna påverkar utfallen av mikroevolutionen. Följande modeller studeras i rapporten:

- Modell 1 - Moranprocess utan selektion
- Modell 2 - Moranprocess med konstant selektion
- Modell 3 - Moranprocess med frekvensberoende selektion
- Modell 4 - Modell för cancertillväxt

Beroende på modell undersöks olika modellparametrar. För den första modellen varieras startfördelningen av genotyperna, liksom den konstanta populationsstorleken. I den andra modellen varieras en selektionskonstant och populationsstorleken. För den tredje varieras värdet på de tre olika parametrarna selektionsintensiteten, vinstvärde och kostnadsvärde i enlighet med Hawk and Dove-spelet. I den fjärde modellen varieras reproduktionstakt, bärformåga samt en specificitetskonstant.

1.3 Avgränsningar

Projektet behöver beakta avgränsningar i modellerna för att förenkla arbetet och analysen. För det första behandlar modellerna endast populationer i en konstant miljö eftersom en varierande miljö skulle kunna påverka selektionen i populationen på oförutsägbara sätt och därmed försvåra modelleringen. För det andra förökar sig organismerna i modellerna endast asexuellt, det vill säga reproduktion endast sker via kloning av sig själva. Detta eftersom en sexuell reproduktion skulle göra det möjligt för olika genkombinationer i avkomman och därmed också försvåra modelleringen. För att förenkla modelleringen ytterligare kommer dessutom populationerna i modellerna endast bestå av två genotyper. Ännu en avgränsning i modellerna är att projektet endast analyserar kortsiktig evolution där jämvikt mellan de olika genotyperna inte hinner uppstå. Dessutom bortses möjligheten till uppkomst av mutationer i populationerna.

1.4 Samhälleliga och etiska aspekter

Då evolutionsprocessen förekommer i flera olika sammanhang på jorden som berör många individer så behöver samhälleliga och etiska aspekter tas hänsyn till vid studier av det. Ett exempel på förekomst av evolution är inom cancertillväxt, där både uppkomst och utbredning av cancerceller kan studeras. Forskning av olika cancerprocesser kan ge kunskap om de mutationer som uppstår och vad de har för fördelaktiga egenskaper. Forskningen av hur cancer ter sig är viktig då det är en vanligt förekommande och dödlig sjukdom som påverkar många.

Evolutionsprocesser beskriver även andra viktiga biologiska företeelser som är av samhällelig eller etisk betydelse, exempelvis uppkomst och utbredning av resistent cancerceller. Detta är av intresse då en stor del av cancers dödlighet inte är avsaknaden av behandling utan cancerens förmåga att utveckla resistens mot behandlingarna. Under cancerbehandling så elimineras de celler som är mottagliga för behandlingen, men då en tumör vanligtvis består av en mångfald olika cancerceller så överlever ofta några celler. Då tumören sedan börjar växa på nytt från de överlevande cancercellerna så besitter den en motståndskraft till den tidigare använda behandlingen [5]. Eftersom den drivande faktorn bakom resistent cancerceller är evolution så är modellering av evolutionsprocesser ett sätt att undersöka detta. Vidare kunskap om resistent cancerceller skulle vara gynnsam för framtagandet av mer effektiv cancerbehandling där resistens kan förhindras eller elimineras. Ett exempel där evolutionsprocessen av cancer undersöks för att förhindra cancercellers utveckling

mot resistens återfinns i artikeln "Cancer Researchers Use Evolution to Target Drug Resistance" [5]. Där utnyttjas vetenskapen av evolution för att erhålla och utveckla bättre behandling av cancer.

Ett annat relevant samband är klimatförändringars påverkan på evolutionsprocesser. Klimatförändringar medför förändrade förhållanden och förutsättningar för många av planetens organismer. Klimatförändringarna bidrar till förändringar i selektionstryck, vilket i sin tur påverkar de arter som lever i de påverkade ekosystemen. Klimatförändringar är därmed ett viktigt område att studera då hela ekosystem av organismer kan rubbas ifall de sätts ur balans. Modeller för evolution kan hjälpa för att få förståelse över hur arterna påverkas av denna förändring, vilket sedan kan användas för bevarandet av mångfalden inom arter och inom ekosystem.

Forskning av evolutionsprocesser kan även vara gynnsamt ur ett försöksperspektiv. Med hjälp av simuleringar kan evolutionära processer åskådliggöras och ge förståelse för hur dessa förlopp ter sig i kroppen. Detta leder till att djur- och människoförsök inte behöver göras för att undersöka samma fenomen. Dock kan inte djur- och människoförsök uteslutas helt eftersom det är ett krav för framställning av mediciner, men behovet kan minskas.

Något som är viktigt att ha i åtanke vid studerande av modeller är att deras tillförlitlighet kan vara bristfällig. Eftersom de verkliga evolutionära processerna är komplexa kan en framtagen modell enkelt misslyckas med att fånga alla aspekter av verkligheten. Ett exempel på detta är att utbredningen av en fördelaktig gen i små populationer bäst beskrivs med stokastiska modeller, medan den i stora populationer beter sig mer deterministiskt. Det är alltså lättare att simulera en liten population, medan det i verkligheten oftare handlar om stora populationer. En otillräcklig tillförlitlighet till modellen kan vidare medföra konsekvenser om avvikelserna från verkligheten inte tas hänsyn till. Även avgränsningarna som gjorts i modellerna, se Avsnitt 1.3, bidrar till en minskad tillförlitlighet. Den ökade förståelsen av evolutionsprocesser är således begränsad och tillförlitligheten bör därför beaktas innan resultaten appliceras på verkligheten.

2 Teori

I följande avsnitt presenteras den evolutionära och matematiska teorin som ligger till grund för de fyra olika modellerna som studerats i projektet. I Bilaga A återfinns en notationslista med notationer som tas upp i följande teori. I Bilaga B återfinns även definitioner och härledningar till teorin och hänvisningar till den kommer ske i detta avsnitt.

2.1 Selektion

Evolution innebär att en population förändras och utvecklas över tid. I denna rapport definieras begreppet evolution som en förändring i förekomsten av specifika gener i en population. Evolutionsförloppet påverkas av många olika faktorer, varav en är selektion. Selektion är en process som grundar sig i att olika typer av organismer i en population konkurrerar med varandra [4]. Bland annat konkurrerar de om vem som har bäst anpassning till sin omgivning, vilket även medför störst chans att överleva och reproducera sig. Genetisk variation, det vill säga varierande genuppsättningar, gör den selektiva processen möjlig och öppnar upp för evolutionär anpassning till omgivningen inom populationen [4].

Uppkomsten av nya genvarianter i en population härstammar bland annat från mutationer som uppstår vid reproduktion. Den selektiva processen påverkar förekomst och frekvens av olika sorters genotyper inom populationen och selektion har sin verkan när de olika genotyperna reproducerar sig i olika takt [4].

Om en population består av två genotyper innebär selektion att den ena genotypen har förbättrat en egenskap som gör den individen bättre lämpad för överlevnad [4]. Ett exempel på detta är ökad reproduktionstakt, där det är mest troligt att den individen med högre reproduktionstakt kommer att ta över populationen. Vid modellering är detta dock inte garanterat i en stokastisk modell eftersom utfallen är slumpmässiga med hänsyn till den beskrivande processen, till skillnad från i en deterministisk modell. En deterministisk modell beskriver populationen i form av differentialekvationer och har ett konstant utfall där en fördel i selektion alltid kommer medföra vinst för den fördelaktiga genotypen [4]. Detta kan redogöras med en deterministisk modell för konstant population med två genotyper, A och B. Frekvensen av genotyp A beskrivs av $x \in [0, 1]$ och frekvensen för B blir därmed $1 - x$. Detta kan beskrivas med en differentialekvation enligt

$$\dot{x} = x(1 - x)(a - b) \quad (2.1)$$

där a och b är reproduktionstakten för genotyp A respektive B. För $a > b$ på intervallet $x \in (0, 1)$ blir $\dot{x} > 0$. Detta innebär att för en population som består av både A- och B-individer och där genotyp A har högre reproduktionstakt, kommer andelen av genotyp A tillslut ta över populationen. Frekvensen av A kommer att konvergera till ett och B konvergerar till noll. Denna deterministiska modell representerar uttrycket "survival of the fittest" [4].

Även i fallet utan selektion kommer utfallet i en stokastisk modell inte kunna etablera samexistens eftersom neutral drift leder till att en av organismtyperna tillslut konkurrerar ut den andra [4]. Konstant selektion definieras som en konstant fördel för den ena genotypen och medför därmed även konstant högre reproduktionstakt.

2.1.1 Frekvensberoende selektion

Selektion kan definieras som frekvensberoende och innebär då att selektionen beror på frekvensen av genotyperna i populationen. Ett exempel på frekvensberoende selektion är frekvensberoende reproduktionstakt, vilket kan modelleras med hjälp av det så kallade Hawk- and Dove-spelet, som är ett spel inom spelteori [4]. Spelet bygger på interaktion inom och mellan arter då det uppstår konflikter kring exempelvis territorium eller mat. De förutbestämda strategierna i detta spel är att hökarna alltid går till attack och duvorna drar sig tillbaka. Konflikterna kan antingen eskalera i strid eller resultera i att den ena, alternativt båda parterna, drar sig tillbaka för att undgå striden.

Striderna värderas i vinstvärdet v , som beskriver förmånen av att vinna en strid, och kostnadsvärdet c , som är kostnaden att förlora en strid.

Vid varje strid i spelet finns en så kallad payoff (jfr. sv. utdelning) [4]. Det finns tre olika typer av strider i Hawk and Dove-spelet. Den första typen av strid är när två hökar möts. Då kommer de båda attackera och endast en hök kan vinna medan den andra kommer att förlora och skadas. Det är 50 procents sannolikhet att vinna eller förlora striden då båda hökarna är lika starka. Detta ger en förväntad payoff för mötet enligt $(v - c)/2$ [4]. För den andra typen av strid är det en hök och en duva som möts. Höken kommer att vinna och mötet får således payoffen v medan duvan drar sig tillbaka och får payoffen noll. För den sista typen av strid är det två duvor som möts, där sannolikheten precis som för två hökar blir 50 procent. En av dem kommer vinna till slut, men ingen av dem kommer att komma till skada, vilket ger mötet en förväntad payoff på $v/2$ [4]. Payoffen för de olika striderna mellan individer kan sammanfattas i en generell payoff-matris [4], som sedan kan skrivas om för Hawk and Dove-spelet enligt

$$\begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} \pi_{1,1} & \pi_{1,2} \\ \pi_{2,1} & \pi_{2,2} \end{pmatrix} \end{array} \Rightarrow \begin{array}{cc} & \begin{array}{cc} H & D \end{array} \\ \begin{array}{c} H \\ D \end{array} & \begin{pmatrix} \frac{v-c}{2} & v \\ 0 & \frac{v}{2} \end{pmatrix}. \end{array} \quad (2.2)$$

Denna payoff påverkar individens reproduktionstakt, vilket gör att bättre strategier bidrar till att individen kan reproducera sig snabbare. De strategier som däremot fungerar sämre gör att de löper större risk att bli utkonkurrerade [4].

I ett spel som beskrivs enligt payoff-matrisen i Ekvation 2.2, kan exempelvis individ A ha en egenskap som är gynnsam i en population då det finns många B och få A, men ogynnsam då det finns många A och få B [4]. Populationen kan då nå en jämvikt mellan antalet A-individer och B-individer. Jämvikter kan delas in i stabila och instabila [4], där en stabil jämvikt innebär att systemet drivs mot jämviktspunkten, medan instabil jämvikt betyder att systemet kommer lämna jämviktssläget om det sätts ur balans. För frekvensberoende selektion inom spelteori kan olika jämviktsslägen studeras beroende på förhållandet mellan spelets payoff [4]. För Hawk and Dove-spelet kan följande fall studeras.

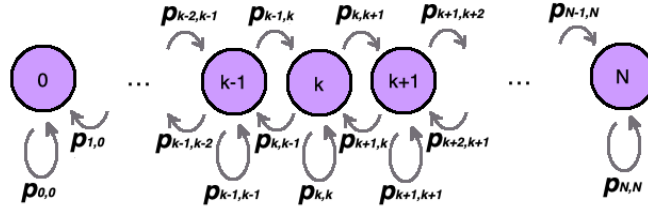
- (i) Dominans, alltså att A dominerar över B, som gäller då $\pi_{1,1} > \pi_{2,1}$ och $\pi_{1,2} > \pi_{2,2}$ [4], vilket innebär att $v > c$. Det finns då ett stabilt jämviktssläge vid A och ett instabilt jämviktssläge vid B. Selektionen kommer att främja A.
- (ii) Samexistens, alltså att A och B existerar stabilt tillsammans, som gäller då $\pi_{1,1} < \pi_{2,1}$ och $\pi_{1,2} > \pi_{2,2}$ [4], vilket innebär att $v < c$. Här finns två instabila jämviktsslägen, då hela populationen består av A respektive B, med ett stabilt jämviktssläge däremellan. Var på intervallet mellan de två instabila jämviktsslägena som den stabila jämviktspunkten, x^* , hamnar kan beräknas med Ekvation 2.3, där x beskriver frekvensen av genotyp A [4].

$$x^* = \frac{\pi_{2,2} - \pi_{1,2}}{\pi_{1,1} - \pi_{1,2} - \pi_{2,1} + \pi_{2,2}} = \frac{\frac{v}{2} - v}{\frac{v-c}{2} - v - 0 + \frac{v}{2}} = \frac{v}{c}. \quad (2.3)$$

2.2 Moranprocess

För att modellera en process som beskriver förändringen av frekvensen av två genotyper i en population, som påverkas av olika typer av selektion, behövs en stokastisk modell. En Moranprocess är den enklaste stokastiska modellen som beskriver dynamiken mellan två genotyper i en population av konstant storlek N [4]. En Moranprocess är en tidsdiskret Markovkedja, vilket betyder att nästa tillstånd endast beror på nuvarande tillstånd och inte på tidigare händelser. För en beskrivning av Markovkedjor se Definition A1. Moranprocessen är en födelse- och dödsprocess, se Definition A2, som beskriver en process där en individ dör och en individ föds i varje tidssteg. Tillstånden består av de möjliga fördelningar mellan antalet A och B i populationen och representeras av antalet

A-individer i . En visualisering av en Moranprocess med dess övergångssannolikheter $p_{i,j}$ finns i Figur 2.1.



Figur 2.1: Illustration av en Moranprocess med övergångssannolikheterna $p_{i,j}$.

I en population med två olika genotyper, A och B, finns det i varje steg i Markovkedjan fyra möjliga händelser beroende på vem som föds respektive dör [4]. De fyra händelserna är följande.

- (i) A dör och A föds, vilket medför att populationens sammansättning av individer kvarstår.
- (ii) B dör och B föds, vilket medför att populationens sammansättning av individer kvarstår, alltså samma utfall som (i).
- (iii) B dör och A föds, vilket medför att antalet individer av typ A ökar.
- (iv) A dör och B föds, vilket medför att antalet individer av typ A minskar.

Dessa fyra händelser genererar tre möjliga utfall för skillnaden i antalet A-individer i populationen. Antingen ökar det med ett, minskar med ett, eller förblir detsamma. Sannolikheten för de olika utfallen kan beskrivas i en övergångsmatrix, $P = [p_{i,j}]$, som är noll-indexerad, där $p_{i,j}$ är sannolikheten att i ett steg gå från tillstånd i till tillstånd j [4]. P är en $(N + 1) \times (N + 1)$ stokastisk matrix, vilket innebär att summan av varje rad är lika med ett. Tillstånden $i = 0$ och $i = N$ är absorberande tillstånd, som betyder att om systemet når ett sådant tillstånd så kan processen inte fortsätta åt något håll, utan är fast i det tillståndet [4]. Matrisen P är tridiagonal vilket innebär att endast $p_{i,i-1}$, $p_{i,i}$ och $p_{i,i+1}$ är nollskilda, för alla heltal i mellan 0 och N .

I Moranprocesserna som undersöks i rapporten är sannolikheten att dö uniformt fördelad över alla individer i populationen. Detta betyder att sannolikheten att en A-individ dör är i/N och motsvarande sannolikhet för B är $(N - i)/N$. Sannolikheterna för reproduktion blir liknande, fast viktat med den relativa reproduktionstakten r för A-individerna [4]. Sannolikheten att en A-individ reproducerar sig blir då $ri/[ri - (N - i)]$ och motsvarande sannolikhet för B blir $(N - i)/[ri - (N - i)]$. Död och reproduktion antas här vara oberoende händelser och därför är sannolikheten att en individ dör och en reproducerar sig i ett steg, produkten av sannolikheterna för de individuella händelserna [4]. Elementen i övergångsmatrisen P kan därmed skrivas

$$\begin{aligned}
 p_{i,i+1} &= \frac{ri}{ri + (N - i)} \frac{N - i}{N}, \\
 p_{i,i-1} &= \frac{N - i}{ri + (N - i)} \frac{i}{N}, \\
 p_{i,i} &= 1 - p_{i,i+1} - p_{i,i-1},
 \end{aligned}
 \tag{2.4}$$

där de absorberande tillstånden är $p_{0,0} = 1$ samt $p_{N,N} = 1$ och alla övriga element är noll [4]. Denna övergångsmatrix beskriver grunden för Modell 1, 2 och 3 som studeras i rapporten där skillnaden beror på hur den relativa reproduktionstakten r definieras.

2.2.1 Moranprocess utan selektion

Den första modellen beskriver en Moranprocess utan selektion. Den relativa reproduktionstakten definieras som $r = 1$, vilket innebär att A och B har samma reproduktionstakt. I detta fall har varken A eller B någon fördel över den andra.

2.2.2 Moranprocess med konstant selektion

Den andra modellen beskriver en process med konstant selektion. För konstant selektion används en selektionskonstant s för att ta hänsyn till genotypernas respektive reproduktionstakt. Här definieras den relativa reproduktionstakten enligt $r = 1 + s$, alltså gynnas A-individer då $s > 0$. För stora värden på s är A-individernas reproduktionstakt kraftigare och då $s = 0$ blir $r = 1$ och fallet reduceras till en Moranprocess utan selektion, se Avsnitt 2.2.1.

2.2.3 Moranprocess med frekvensberoende selektion

Som tidigare nämnts i Avsnitt 2.1.1 innebär frekvensberoende selektion att individernas reproduktionstakt beror på frekvensen av varje genotyp i populationen. I en population med två strategier, A och B, interagerar individerna slumpmässigt med varandra och då två individer möts agerar de utifrån sina förutbestämda strategier [4]. Sannolikheten för en A-individ att möta en A-individ är $(i - 1)/(N - 1)$ eftersom det finns $(i - 1)$ A-individer av totalt $(N - 1)$ individer i populationen som den hade kunnat möta. Enligt samma resonemang blir därför sannolikheten för en A att möta en B $(N - i)/(N - 1)$, sannolikheten att en B möter en A blir $i/(N - 1)$ och sannolikheten att en B möter en B $(N - i - 1)/(N - 1)$.

Utifrån den generella payoff-matris beskriven i Ekvation 2.2 kan den förväntade payoffen F för vardera genotyp definieras samt skrivas om för Hawk- and Dove-spelet [4]. För A respektive B blir därmed den förväntade payoffen följande

$$F_i^A = \frac{\pi_{1,1}(i - 1) + \pi_{1,2}(N - i)}{N - 1} = \frac{\frac{v-c}{2}(i - 1) + v(N - i)}{N - 1}, \quad (2.5)$$

$$F_i^B = \frac{\pi_{2,1}i + \pi_{2,2}(N - i - 1)}{N - 1} = \frac{\frac{v}{2}(N - i - 1)}{N - 1}.$$

Reproduktionstakterna r_i^A och r_i^B för A respektive B beror på deras förväntade payoff F [4] och kan uttryckas som

$$r_i^A = 1 - w + wF_i^A, \quad r_i^B = 1 - w + wF_i^B, \quad (2.6)$$

där w är selektionsintensiteten och definieras på intervallet $w \in [0, 1]$ [4]. När $w = 0$ påverkas inte reproduktionstakten av spelet, vilket innebär att de två strategierna är neutrala. När $w = 1$ är selektionen stark, vilket innebär att reproduktionstakten endast beror på den förväntade payoffen F [4]. Den relativa reproduktionstakten r då vi har i A-individer blir därmed

$$r = \frac{r_i^A}{r_i^B} = \frac{1 - w + wF_i^A}{1 - w + wF_i^B}. \quad (2.7)$$

2.3 Selektion hos cancerceller

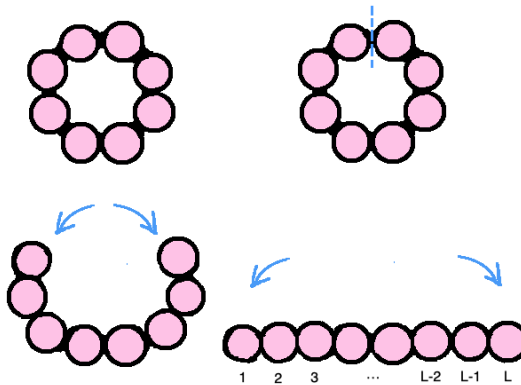
Som tidigare nämnts innebär selektion att organismer i en population konkurrerar med varandra. Detta är precis vad som sker när cancerceller konkurrerar ut kroppens friska celler. Cancercellers främsta egenskap är att de förökar sig i en snabbare takt än friska celler och på så sätt utvecklar sjukdom [6]. På senare tid har det dock framkommit att det inte bara är deras reproduktionstakt som är avgörande, utan även deras bähförmåga har en påverkan på selektionen. En ökad bähförmåga innebär att cellerna uppvisar en högre celltäthet, vilket har uppstått genom att de har anpassat sin omgivning till sin fördel för att kunna växa till sig [6]. Sådan anpassning kan göras genom exempelvis angiogenes, vilket innebär att cancercellerna utsöndrar ämnen som stimulerar tillväxten av nya blodkärl [7]. Detta leder till att celler kan växa på fler ställen och med högre täthet än de borde, vilket ger en ökad bähförmåga. Ett annat sätt som cancercellerna ökar sin bähförmåga på är genom att utsöndra autokriner signaler som förbättrar deras förmåga att genomgå celledelning i hög celltäthet [6]. På grund av att omgivningen är förändrad runt en cell med ökad bähförmåga så kan de angränsande cellerna dra nytta av denna [6]. Dock är det specificiteten på modifieringen

av omgivningen som avgör hur mycket av bärformågan som delas. Om det är en väldigt specifik anpassning för cancercellen är det svårare för friska celler att utnyttja det [6].

Bröstcancer bildas i bröstkanalerna som är beklädda med ett lager av epitelceller [6]. När en mutation sker i en av dessa epitelceller, så att den blir cancerogen, börjar den konkurrera med epitelcellerna. Cancercellerna, som fördelaktligen har både snabbare reproduktionstakt och en högre bärformåga, kan därför både föröka sig snabbt samt växa på varandra. Detta leder till att hela kanalen tillslut täpps igen av cancerceller. När de sedan blir tillräckligt många kan kanalen spricka och canceren kan då sprida sig till omkringliggande vävnad och således bli invasiv [6].

2.3.1 Modell för concertillväxt

Vid modellering av bröstcancer kan tvärsnittet av en bröstkanal studeras. Detta kan ses som en cirkel beklädd med ett lager av epitelceller, där varje cell sitter på varsin plats, även kallad gitterpunkt [6]. Denna cirkel kan sedan vecklas ut och istället ses som en rad med platser där celler kan befinna sig, vilket visas i Figur 2.2, där första och sista cellen i raden betecknas som att de egentligen sitter bredvid varandra. På så sätt förenklas modelleringen av problemet till endast en dimension. På platserna kan cellerna antingen dela sig eller dö och för varje plats kan det antingen befinna sig en epitelcell, ingen cell, eller en eller flera cancerceller [6].



Figur 2.2: Illustration av hur ett tvärsnitt av bröstkanalen kan ses som en rad gitterpunkter.

Specificiteten, γ , anger hur specifik modifieringen är för cellen och på så sätt hur bärformågan kan fördelas [6]. Den varierar på intervallet $[0,1]$, där $\gamma = 0$ innebär att all bärformåga fördelas jämnt över de angränsande cellerna medan för $\gamma = 1$ fördelas den inte alls utan behålls till den enskilda cellen. Eftersom celler kan dela med sig av sin bärformåga så finns även en så kallad effektiv bärformåga, det vill säga den bärformåga som celler faktiskt upplever [6]. Den kan beskrivas av en ekvation enligt

$$K_{i,l} = \gamma \cdot k_i + (1 - \gamma) \cdot \bar{K}_l \quad (2.8)$$

där k_i är den inre bärformågan hos en cell, och \bar{K}_l är medelvärdet av k_i för celler på plats l [6]. Detta medelvärde kan beräknas enligt

$$\bar{K}_l = \frac{1}{N_l} \sum_{i=1}^{N_l} k_i \quad (2.9)$$

där N_l är antalet celler på plats l [6].

Huruvida en cell delar sig, var dottercellen sätter sig vid delning och om cellerna överlever sker med vissa sannolikheter [6]. Sannolikheten för celledelning är $\rho_i^r = r_i \Delta t$, där r_i är reproduktionstakten för cell i och Δt är tidsstegen. Om cellen delar sig så hamnar dottercellen på samma plats som modercellen med en sannolikhet på $\rho_{i,l} = 1 - N_l / K_{i,l}$. Om det inte sker så kan dottercellen antingen

hamna på platsen $l+1$ eller $l-1$ med en sannolikhet på $\rho_{i,l+1} = 1 - N_{l+1}/K_{i,l+1}$ respektive $\rho_{i,l-1} = 1 - N_{l-1}/K_{i,l-1}$. Sker inget av dessa utfall så dör dottercellen. Huruvida modercellen överlever efter celledelning bestäms av sannolikheten $\rho_i^a = \delta\Delta t$ där δ är avdöningstakten. Sannolikheterna för den stokastiska processen kan användas för att undersöka hur väl cancercellerna kan ta över bröstkanalen och konkurrera ut de friska cellerna.

2.4 Absorptionssannolikhet

Sannolikheten att en Markovprocess fastnar i ett absorberande läge givet ett visst initialtillstånd kallas för absorptionssannolikhet [4]. Vanligen betraktas absorptionssannolikhet gällande en av genotyperna, vilket givet ett startantal vid tidsstart anger sannolikheten att populationen endast kommer bestå av den genotypen i framtiden [4].

Låt P_{a_i} beteckna absorptionssannolikheten som är sannolikheten att nå tillstånd N , det vill säga att hela populationen består av A-individer, utifrån starttillståndet i [4]. Den generella formeln för absorptionssannolikheten ges då av

$$P_{a_i} = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^j p_{k,k-1}/p_{k,k+1}}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j p_{k,k-1}/p_{k,k+1}}. \quad (2.10)$$

Härledning av uttrycket för P_{a_i} återfinns i Avsnitt B.3. Fixeringssannolikhet, P_f , är ett specialfall av absorptionssannolikhet och innebär att endast starttillstånd $i = 1$ beaktas [4]. Fenomenet kan liknas med en population av enbart B där en A-individ producerats genom exempelvis mutation eller migration.

2.4.1 Absorptionssannolikhet utan selektion

Absorptionssannolikheten för en Moranprocess utan selektion kan härledas från den generella formeln för absorptionssannolikhet i Ekvation 2.10, där $p_{i,i-1} = p_{i,i+1}$. Det leder till följande analytiska uttryck

$$P_{a_i} = \frac{i}{N} \quad (2.11)$$

där i är starttillståndet i populationen och N är totala antalet i populationen [4].

Fixeringssannolikheten ges därmed av följande uttryck

$$P_f = \frac{1}{N}. \quad (2.12)$$

2.4.2 Absorptionssannolikhet med konstant selektion

För en Moranprocess med konstant selektion kan följande uttryck definieras från förenkling av övergångsmatrisen definierad i Ekvation 2.4 med relativ reproduktionstakt, $r = 1 + s$, enligt

$$\frac{p_{i,i-1}}{p_{i,i+1}} = \frac{1}{1+s}. \quad (2.13)$$

Absorptionssannolikheten kan därmed härledas genom den generella formel i Ekvation 2.10, till följande analytiska uttryck

$$P_{a_i} = \frac{1 - 1/(1+s)^i}{1 - 1/(1+s)^N} \quad (2.14)$$

där s är selektionskonstanten, i starttillståndet i populationen och N totala antalet i populationen [4].

Fixeringssannolikheten ges därmed analytiskt av följande uttryck

$$P_f = \frac{1 - 1/(1+s)}{1 - 1/(1+s)^N} \quad (2.15)$$

som även kan skrivas om till

$$P_f = \frac{(1+s)^N - (1+s)^{N-1}}{(1+s)^N - 1}. \quad (2.16)$$

Då $s = 0$ fås fallet med ingen selektion, vilket skall medföra att fixeringssannolikheten $P_f = 1/N$ enligt Ekvation 2.12. Detta kan visas med hjälp av gränsvärde då $s \rightarrow 0$,

$$\lim_{s \rightarrow 0} \frac{1 - 1/(1+s)}{1 - 1/(1+s)^N} \quad (2.17)$$

och genom att tillämpa l'Hôpitals regel blir uttrycket

$$\lim_{s \rightarrow 0} \frac{1/(1+s)^2}{N(1+s)^{-N-1}} = \frac{1}{N}. \quad (2.18)$$

2.4.3 Absorptionssannolikhet med frekvensberoende selektion

Absorptionssannolikheten för en Moranprocess med frekvensberoende selektion kan beräknas med hjälp av den generella formeln för absorptionssannolikhet från Ekvation 2.10 [4]. Från uttrycket för övergångsmatrisen, definierad i Ekvation 2.4, med relativ reproduktionstakt för frekvensberoende selektion enligt $r = r_i^A/r_i^B$, kan följande omskrivning göras

$$\frac{p_{i,i-1}}{p_{i,i+1}} = \frac{r_i^B}{r_i^A}. \quad (2.19)$$

Absorptionssannolikheten kan därmed beskrivas enligt

$$P_{a_i} = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^j r_k^B / r_k^A}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j r_k^B / r_k^A}. \quad (2.20)$$

Fixeringssannolikheten kan därför numeriskt beskrivas med följande ekvation

$$P_f = 1 / \left(1 + \sum_{k=1}^{N-1} \prod_{i=1}^k \frac{r_i^B}{r_i^A} \right). \quad (2.21)$$

3 Metod

I detta avsnitt redovisas metoden som användes i projektet. Modellerna beskrivs utifrån två genotyper, A och B, där jämförandet utgår från A-individerna. Genotyp A kan därmed ses som cancerceller medan genotyp B kan ses som de friska cellerna. All kod återfinns i Bilaga C och kommer att hänvisas till i detta avsnitt.

3.1 Modellering av Moranprocessen

De tre första modellerna är varianter på Moranprocessen. För varje steg i Moranprocessen avgörs utfallet utifrån övergångssannolikheterna för de olika modellerna, se Avsnitt 2.2. Övergångssannolikheterna är det enda som särskiljer de tre första modellerna från varandra och därför beskrivs först metoden som använts för att simulera en allmän Moranprocess.

Koden simulerade Moranprocessen som en serie av steg. I varje steg i Markovkedjan beräknades övergångssannolikheterna och utfallet slumpades fram. Detta upprepades tills fixering uppnåddes. Varje process simulerades 1000 gånger för varje parameterintervall.

Ett 95 % tvåsidigt konfidensintervall för absorptionssannolikheten respektive fixeringssannolikheten beräknades enligt Avsnitt B.2 beroende på vilka parametervärden som användes. Därefter plottades resultaten för Modell 1 och 2 tillsammans med sin analytiska lösning, och Modell 3 plottades tillsammans med sin numeriska lösning.

3.1.1 Modell 1 - Moranprocess utan selektion

För Modell 1 undersöktes absorptionssannolikheten $P_a = P_a(N_0, N)$ för totalpopulationerna $N = 10$ och $N = 100$ med nio olika startvärden N_0 sådana att

$$\frac{N_0}{N} = \{1/10, 2/10, 3/10, \dots, 9/10\}. \quad (3.1)$$

Övergångssannolikheterna som användes för Modell 1 är givna i Ekvation 2.4, där den relativa reproduktionstakten är $r = 1$. Den analytiska lösningen för absorptionssannolikheten är given av Ekvation 2.11. Koden till Modell 1 återfinns i Avsnitt C.1.

3.1.2 Modell 2 - Moranprocess med konstant selektion

För Modell 2 undersöktes fixeringssannolikheten $P_f = P_f(N_0, N, s)$ för totalpopulationerna $N = 10$ och $N = 100$ med en varierande selektionskonstant på $s = \{0, 0.5, \dots, 5.5\}$. Övergångssannolikheterna som användes är givna i Ekvation 2.4 där den relativa reproduktionstakten är $r = 1 + s$ och den analytiska lösning för fixeringssannolikheten som användes är given av Ekvation 2.15. Koden för Modell 2 finns i Avsnitt C.2.

3.1.3 Modell 3 - Moranprocess med frekvensberoende selektion

För Modell 3 undersöktes fixeringssannolikheten $P_f = P_f(N_0, N, w, v, c)$ för en totalpopulation på $N = 100$ samt selektionsintensiteterna $w = 0.1$ och $w = 0.5$. Detta studerades för de två olika fallen dominans, när $v > c$, och samexistens, när $v < c$, i enlighet med Avsnitt 2.1.1. För samexistens i Hawk and Dove-spelet finns det begränsningar på v och c sådana att $v/c \in [0, 1]$. Därmed sattes kostnadsvärdet $c = 1$ och vinstvärdet varierades enligt $v = \{0.1, 0.2, \dots, 0.8, 0.9\}$. För dominans varierades istället vinstvärdet enligt $v = \{1, 1.5, \dots, 4.5, 5\}$.

Övergångssannolikheterna som användes för Modell 3 är givna i Ekvation 2.4 där den relativa reproduktionstakten är $r = r_i^A/r_i^B$. Fixeringssannolikhetens numeriska lösning är given av Ekvation 2.21. Koden för Modell 3 finns i Avsnitt C.3.

3.2 Modell 4 - Modell för cancertillväxt

För Modell 4 skapades en kod baserad på teorin om bröstcancer cellers beteende, se Avsnitt 2.3 och 2.3.1. Modelleringen delades upp för de olika genotyperna för att särskilja dem åt men sågs som en och samma process på samma tvärsnitt i en bröstkanal. Därmed skapades en lista för A och en lista för B, båda med L antal gitterpunkter, där i i varje punkt motsvarar antalet celler på den platsen. Lista B innehöll från början en cell per gitterpunkt medan Lista A innehöll noll. Sedan slumpades det fram en gitterpunkt där en B byttes ut mot en A. Alltså ändrades platsen i Lista B till noll och i Lista A till ett.

I varje tidssteg i processen slumpades det fram för varje cell i båda listorna huruvida cellen delar sig eller inte, enligt sannolikheten för celledning, se Avsnitt 2.3.1. Om cellen delar sig så kan dottercellen antingen stanna på samma plats som modercellen eller placeras på en grannplats till höger eller vänster i listan. Detta slumpades också fram utifrån förflyttningssannolikheterna, se Avsnitt 2.3.1. Om dottercellen ej hade möjlighet att placeras på någon av dessa tre platser så dog den. I varje tidssteg finns det även en risk att modercellen dör, se Avsnitt 2.3.1.

För modellen undersöktes fixeringssannolikheten $P_f = P_f(k_A, r_A, k_B, r_B, \gamma)$. Varje process utfördes 1000 gånger för varje varierande värde och därefter beräknades ett 95 % tvåsidigt konfidensintervall enligt Avsnitt B.2 för fixeringssannolikheten.

Parametrar för inre bärformåga och reproduktionstakt på genotyp B var konstanta enligt $k_B = 1$ och $r_B = 1$. Även avdöningstakten och tidsstegets längd var konstanta parametrar och sattes till $\delta = 0.1$ respektive $\Delta t = 0.01$. Parametrarna som varierades var den inre bärformågan k_A och reproduktionstakten r_A för genotyp A, samt specificiteten γ . Den inre bärformågan varierades enligt $k_A = \{1, 2, 3, 4, 5\}$, reproduktionstakten varierades enligt $r_A = \{1, 1.2, 1.4, 1.6, 1.8, 2\}$ och specificiteten varierades enligt $\gamma = \{0, 0.25, 0.5, 0.75, 1\}$. Fixeringssannolikheten undersöktes för $L = 10$ och $L = 100$ och visualiserades som både grafer och färgdiagram. Koden kan studeras i Avsnitt C.4.

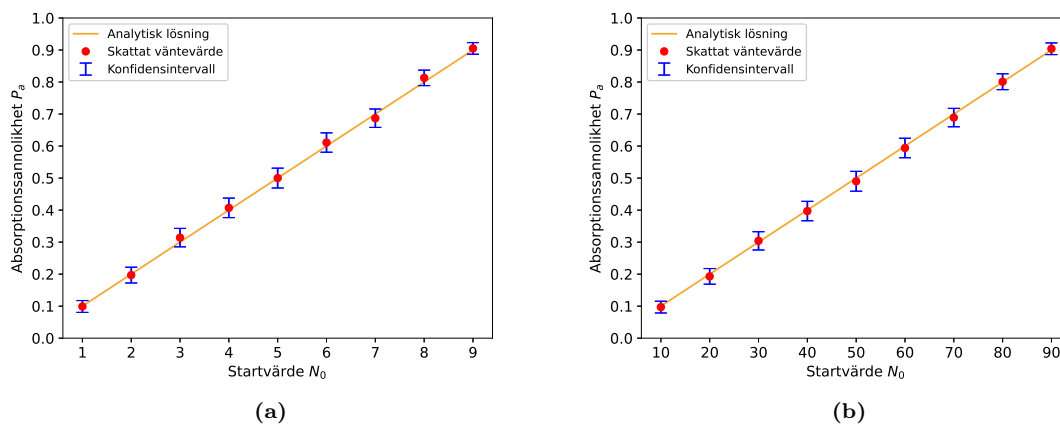
4 Resultat

I detta avsnitt presenteras resultatet som erhållits från simulering av de fyra olika modellerna. Modellernas resultat presenteras i turordning från Modell 1 till Modell 4. Rådata för resultatet finns samlat i Bilaga D och kommer hänvisas till i detta avsnitt.

4.1 Modell 1 - Moranprocess utan selektion

Resultatet för Modell 1 innefattar 1000 simuleringar av varje värde av kvoten $N_0/N = \{0.1, 0.2, \dots, 0.9\}$ för de två olika populationsstorlekarna, $N = 10$ och $N = 100$. Varje simulering ger antingen vinst för A eller vinst för B beroende på vilken genotyp som fixerats i den omgången. De beräknade 95 % tvåsidiga konfidensintervallen för absorptionssannolikheterna återfinns i Tabell D.1.

Det sammanställda resultatet för de stokastiska simuleringarna kan skådas i Figur 4.1. Den analytiska lösningen för Modell 1 togs fram med Ekvation 2.11 och kan ses tillsammans med de stokastiska simuleringarna i figuren.

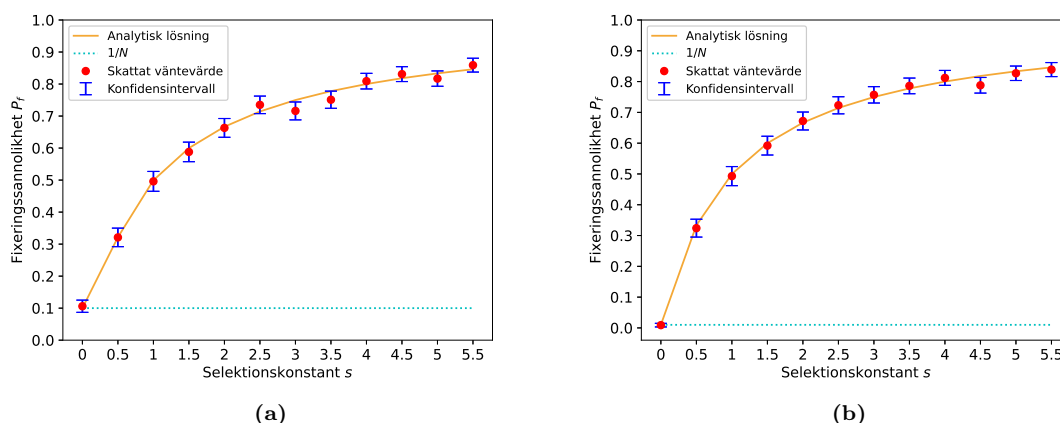


Figur 4.1: Visualisering av skattat väntevärde samt 95 % konfidensintervall av absorptionssannolikhet för Modell 1 för varierande startvärden N_0 för populationsstorlekarna (a) $N = 10$ och (b) $N = 100$. Den analytiska lösningen är markerad i brandgult och följer Ekvation 2.11. Konfidensintervall är markerade i blått och skattat väntevärde är markerade i rött.

Från simuleringarna av de olika populationsstorlekarna och det varierande startvärdet N_0 erhöles att samma värde av kvoten N_0/N korresponderar med absorptionssannolikheten P_a , enligt Ekvation 2.11. Detta medför att det är förhållandet mellan startvärdet och populationsstorleken som påverkar absorptionssannolikheten i Modell 1. Det kan även utläsas att absorptionssannolikheten växer linjärt som funktion av N_0 . Den analytiska lösningen för Modell 1, $y = N_0/N$, beskriver samma förhållande mellan startvärde och populationsstorlek som utläses av de utförda stokastiska simuleringarna.

4.2 Modell 2 - Moranprocess med konstant selektion

För Modell 2 utfördes 1000 simuleringar med varierande selektionskonstant enligt $s = \{0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5\}$ för att studera fixeringssannolikheten för populationsstorlekarna $N = 10$ och $N = 100$. Skattat väntevärde samt beräknade tvåsidiga 95 % konfidensintervall för fixeringssannolikhet återfinns i Tabell D.2. Resultatet för de stokastiska simuleringarna kan ses i Figur 4.2 där även den analytiska lösningen för Modell 2, enligt Ekvation 2.15, är illustrerad.



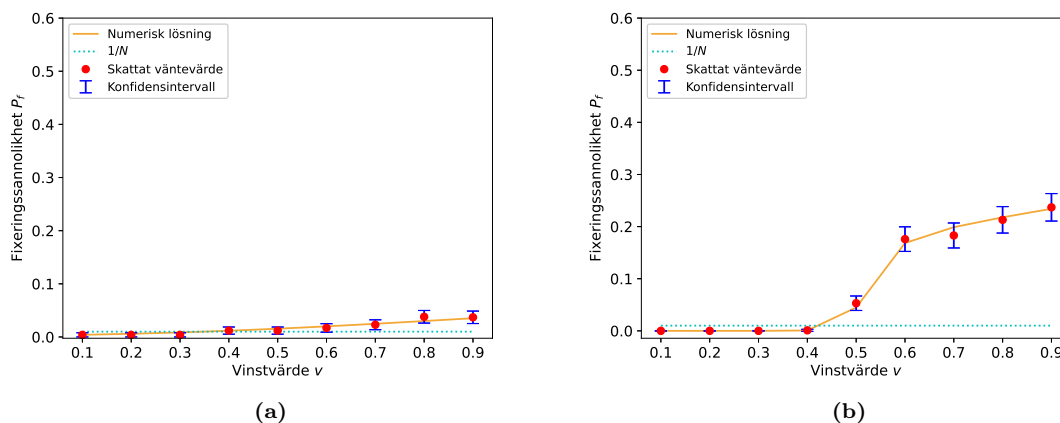
Figur 4.2: Visualisering av skattat väntevärde samt 95 % konfidensintervall av fixeringssannolikheten för Modell 2 för varierande s -värden för populationsstorlekarna (a) $N = 10$ och (b) $N = 100$. Den analytiska lösningen är markerad i brandgult och följer Ekvation 2.15. Konfidensintervall är markerade i blått och skattat väntevärde är markerade i rött. Fixeringssannolikhet för det neutrala fallet utan selektion visas med den streckade linjen.

Resultatet i Figur 4.2 visar att de stokastiska simuleringarna stämmer överens med den analytiska lösningen för fixeringssannolikheten av Modell 2. Figuren visar även att fixeringssannolikheten blir högre med ökande värden på selektionskonstanten.

4.3 Modell 3 - Moranprocess med frekvensberoende selektion

För Modell 3 studerades fixeringssannolikhet för två fall av jämviktpunkter. Det första fallet som studerades var samexistens. Då gjordes 1000 simuleringar med varierande vinstvärden enligt $v = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ och med en totalpopulation på $N = 100$ för att studera fixeringssannolikheten för selektionsintensiteter på $w = 0.1$ och $w = 0.5$. De beräknade tvåsidiga 95 % konfidensintervallen för fixeringssannolikheten återfinns i Tabell D.3.

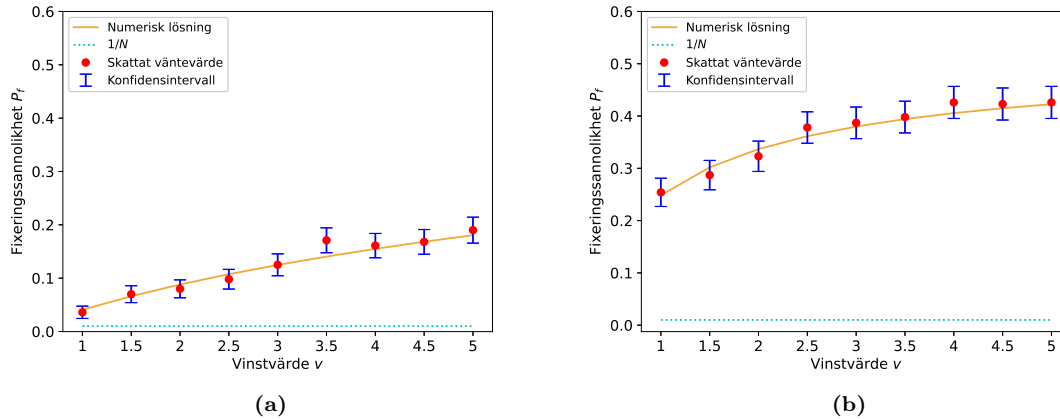
Resultatet för fixeringssannolikheten vid samexistens kan skådas i Figur 4.3. Observera att figuren endast visar värden av fixeringssannolikheten mellan $P_f = [0, 0.6]$. Fixeringssannolikhetens numeriska lösning för Modell 3 uttrycks enligt Ekvation 2.21.



Figur 4.3: Visualisering av skattat väntevärde samt 95 % konfidensintervall för fixeringssannolikheten av Modell 3 vid samexistens för olika selektionsintensiteter och varierande vinstvärden då $N = 100$. Den numeriska lösningen är markerad i brandgult och följer Ekvation 2.21. De olika selektionsintensiteter som studeras är (a) $w = 0.1$ och (b) $w = 0.5$. Konfidensintervall är markerade i blått och skattat väntevärde är markerade i rött. Fixeringssannolikhet för det neutrala fallet utan selektion visas med den streckade linjen.

För det andra fallet med jämviktspunkter dominans, när A dominerar i populationen, studerades fixeringssannolikheten också för selektionsintensiteter på $w = 0.1$ och $w = 0.5$. Det utfördes 1000 simuleringar med varierande vinstvärden enligt $v = \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ för en totalpopulation på $N = 100$. De beräknade tvåsidiga 95 % konfidensintervallen för fixeringssannolikheterna finns i Tabell D.4.

I Figur 4.4 presenteras resultatet för fixeringssannolikhet vid dominans. Observera att figuren endast visar värden av fixeringssannolikheten mellan $P_f = [0, 0.6]$. Fixeringssannolikhetens numeriska lösning uttrycks enligt Ekvation 2.21.



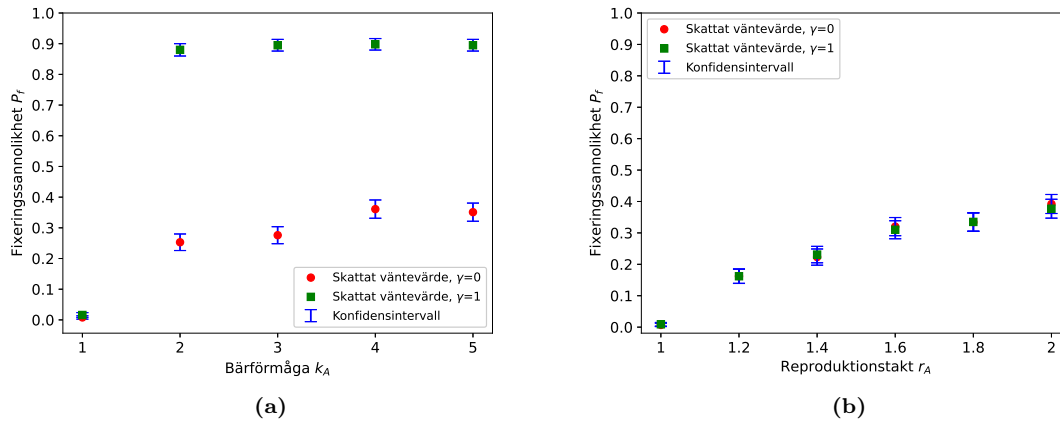
Figur 4.4: Visualisering av skattat väntevärde samt 95 % konfidensintervall av fixeringssannolikhet för Modell 3 vid dominans då $N = 100$ samt varierande vinstvärden. Detta för olika selektionsintensiteter enligt (a) $w = 0.1$, (b) $w = 0.5$. Dessa är plottade tillsammans med den numeriska lösningen som är markerad i brandgult och följer Ekvation 2.21. Konfidensintervall är markerade i blått och skattat väntevärde är markerade i rött. Fixeringssannolikhet för det neutrala fallet utan selektion visas med den streckade linjen.

Figur 4.3 och 4.4 visar att de stokastiska simuleringarna stämmer överens med de plottade numeriska lösningarna. Fixeringssannolikheten ökar med en ökning av vinstvärdet v . Det kan även ses att vid högre värden på w så är fixeringssannolikheten större.

4.4 Modell 4 - Modell för concertillväxt

För Modell 4 sattes varierande värde av bärformågan för A-individerna enligt $k_A = \{1, 2, 3, 4, 5\}$ och reproduktionstakten enligt $r_A = \{1, 1.2, 1.4, 1.6, 1.8, 2\}$. De satta parametervärdena var $\Delta t = 0.01$ och $\delta = 0.1$.

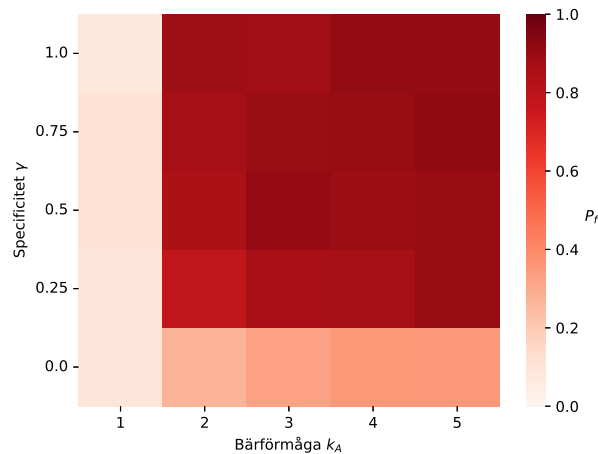
Resultatet för hur fixeringssannolikheten påverkas av bärformåga och reproduktionstakt med olika specificiteter då $L = 100$ kan ses i Figur 4.5. Samma resultat fast för $L = 10$ återfinns i Figur D.1.



Figur 4.5: Illustration av skattat väntevärde samt 95 % konfidensintervall av fixeringssannolikhet för cancertillväxt när $L = 100$. Detta plottat för specificiteterna $\gamma = 0$, som är de röda punkterna, och $\gamma = 1$, som är de gröna punkterna. Parametervärdena antar (a) $r_A = 1$ och varierande k_A , samt (b) $k_A = 1$ och varierande r_A .

Från Figur 4.5 framgår det tydligt att specificitetkonstanten har stor påverkan då genotyp A har övertag med sin bärförmåga, men att den inte påverkar då genotyp A har en ökad reproduktionstakt. Det kan även skådas att fixeringssannolikheten ökar både då värdena på bärförmågan och reproduktionstakten ökar. Vid jämförelse mellan Figur 4.5 och D.1 framgår det att antalet gitterpunkter inte påverkar fixeringssannolikheten med undantag för första punkten i samtliga figurer.

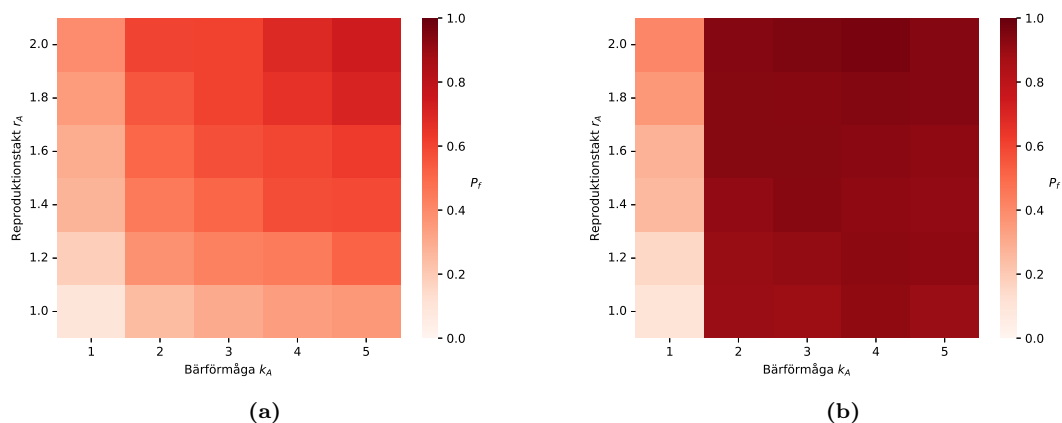
I Figur 4.6 presenteras resultatet för fixeringssannolikhet för cancertillväxt då värden på specificiteten γ och bärförmågan k_A varieras enligt $\gamma = \{0, 0.25, 0.5, 0.75, 1\}$ respektive $k_A = \{1, 2, 3, 4, 5\}$.



Figur 4.6: Illustration av värden på fixeringssannolikheter för cancertillväxt för $L = 10$ med varierande värden på specificitetskonstanten γ och bärförmågan för genotyp A, k_A .

I Figur 4.6 syns det att vid värden då $\gamma > 0$ och $k_A > 1$ blir fixeringssannolikheten väldigt stor. Då $k_A = 1$ är fixeringssannolikheten låg för alla värden på specificiteten γ . För $\gamma = 0$ kan det observeras att fixeringssannolikheten ökar något med ökande värde på bärförmågan, k_A .

I Figur 4.7 återfinns två färgdiagram som beskriver hur fixeringssannolikheten påverkas då både bärförmåga och reproduktionstakt varieras samtidigt. En mörkare röd färg innebär en större fixeringssannolikhet. Bärförmågan och reproduktionstakten varieras enligt $k_A = \{1, 2, 3, 4, 5\}$ respektive $r_A = \{1, 1.2, 1.4, 1.6, 1.8, 2\}$.



Figur 4.7: Illustration av fixeringssannolikhet för cancertillväxt för $L = 10$ med varierande värden på reproduktionstakten för A, r_A , och bärförmågan för A, k_A . Värdet på specificitetskonstanten γ antar (a) $\gamma = 0$ och (b) $\gamma = 1$.

Det finns en tydlig skillnad i Figur 4.7 på hur stor fixeringssannolikheten blir mellan $\gamma = 0$ och $\gamma = 1$ då både bärförmåga och reproduktionstakt varierar samtidigt. För $\gamma = 0$ observeras en gradient där sannolikheten för fixering är låg för låga värden på r_A och k_A , till att bli en högre sannolikhet för höga värden på r_A och k_A . För $\gamma = 1$ har k_A en större påverkan på utfallet än r_A då fixeringssannolikheten direkt blir väldigt hög när k_A blir större än ett.

5 Diskussion

Projektet undersökte evolution på mikroskala med fyra stokastiska modeller. Modellerna analyserade hur olika modellspecifika parametrar påverkar evolutionens utfall genom att uppskatta absorptions- och fixeringssannolikhet. De tre första modellerna som undersöktes var baserade på en stokastisk Moranprocess där olika typer av selektion studerades, medan den fjärde modellen var baserad på en modell av bröstcancer. I följande avsnitt presenteras en diskussion av erhållet resultat för dessa modeller och hur det kan kopplas till verkligheten. Projektets metoder och begränsningar kommer också utvärderas.

5.1 Diskussion av resultat

Från resultatet framgår vikten av att simulera stokastiskt till skillnad från deterministiskt i studie av mindre populationer. En rent deterministisk modell har ingen inverkan av slumpen och skulle därmed alltid ge samma resultat. Om en individ i populationen har en fördelaktig egenskap gentemot alla andra individer så skulle den egenskapen i en deterministisk modell på så vis alltid tillsluta sig över populationen. Detta är inte fallet för stokastiska processer vilket framgår tydligt i Figur 4.1 som visar en Moranprocess utan selektion. Där konkurrerar inte individerna med varandra, eftersom det inte finns något övertag, men en genotyp fixeras ändå alltid i slutändan enbart genom neutral drift. Detta stämmer bättre överens med hur evolution beter sig under kort tid och i små populationer.

Resultatet för de olika Moranprocesserna illustrerar att så fort en av genotyperna har en högre reproduktionstakt fås en större fixeringssannolikhet jämfört med det neutrala fallet, då $P_f = 1/N$. För den andra modellen, *Moranprocess med konstant selektion*, observeras i Figur 4.2 att då selektionskonstanten s är större än 0 så är fixeringssannolikheten större än det neutrala fallet, vilket betyder att fixering av A-individen främjas. För den tredje modellen, *Moranprocess med frekvensberoende selektion*, i fallet där hökar och duvor lever i samexistens, illustrerar Figur 4.3 att då vinstvärdet blir tillräckligt stort så slår fixering av hökar igenom och fixeringssannolikheten blir större än det neutrala fallet $1/N$. I fallet av frekvensberoende selektion där hökar är dominanta illustrerar Figur 4.4 att precis som i fallet av konstant selektion, när selektionskonstanten s är större än 0, så är fixeringssannolikheten större än det neutrala fallet.

För den fjärde modellen, *Modell för cancercellväxt*, visar resultatet tydligt att specificitetskonstanten γ har stor påverkan när A-individerna har fördel i bärformåga gentemot B-individerna, se Figur 4.5a. För $\gamma = 0$ observeras en ökning av fixeringssannolikheten då A-individernas bärformåga ökas, och för $\gamma = 1$ antas fixering av genotyp A kring 90 procent av gångerna. Dock påverkar specificitetskonstanten inte fixeringssannolikheten då A-individerna har en ökad reproduktionstakt, se Figur 4.5b. Detta stämmer överens med teorin om de olika sannolikheterna för hur celler fördelas i en population, det vill säga vilken gitterpunkt dottercellerna hamnar på, se Avsnitt 2.3.1. I Ekvation 2.8, som beskriver den effektiva bärformågan, finns det två termer, där en av dem försvinner för extremvärden på γ . Med andra ord, när $\gamma = 0$ fås $K_{i,l} = \bar{K}_l$, och när $\gamma = 1$ får $K_{i,l} = k_i$. Eftersom medelvärdet \bar{K}_l , även kan innehålla en B-individs bärformåga som sänker dess värde, ger det att $\bar{K}_l \leq k_i$. Vidare används dessa värden i sannolikheten $\rho_{i,l} = 1 - N_l/K_{i,l}$. Störst sannolikhet ges därmed vid störst värde på $K_{i,l}$, det vill säga för $\gamma = 1$. Samma resonemang gäller för sannolikheterna att dottercellen hamnar på platserna bredvid. Sannolikheten att dottercellen dör blir då även mindre för $\gamma = 1$ eftersom sannolikheten är större att hamna på en av de tre platserna. För reproduktionstakten är det rimligt att resultaten blir densamma för $\gamma = 0$ som för $\gamma = 1$ eftersom γ endast påverkar hur mycket bärformåga som delas med omgivningen och har alltså ingen påverkan på reproduktionstakten. Detta kan även förklaras matematiskt då r_A inte ingår i Ekvation 2.8.

Att det blir en ökning av fixeringssannolikheten då bärformågan ökar när $\gamma = 0$, tyder på att egenskapen att kunna bära flera celler på samma plats är fördelaktig. Dessutom innebär det att ju fler celler som kan få plats, desto större fördel tillkommer till cellen. När $\gamma = 1$ fås en ännu större ökning av fixeringssannolikheten direkt då bärformågan blir större. Detta betyder att egenskapen att dessutom kunna behålla bärformågan till sig själv gynnar cellen mer än om den skulle dela

med sig av sin bärformåga. Resultatet är därmed rimligt eftersom det innebär att om bärformågan skulle delas med omgivande celler så skulle en B-individ, med låg bärformåga, kunna ta del av en ökad bärformåga från en A-individ. B-individen skulle då också kunna växa till sig och således motverka fixering av genotyp A. Detta fenomen kan vidare studeras i Figur 4.6 som visualiserar hur fixeringssannolikheten påverkas då både specificiteten och bärformågan varieras samtidigt. I figuren framgår det en distinkt skillnad på fixeringssannolikheten direkt då specificiteten är större än 0. Detta tyder på att egenskapen att kunna behålla sin egen bärformåga, om så bara lite, ger en stor fördel även om den också delas med omgivningen. Från detta resultat var det därför inte nödvändigt att studera varierande specificitet mer utförligt i projektet. För cancerutveckling innebär resultatet att så fort en cancercell muteras så att den kan behålla sin bärformåga, om bara lite grann, så ger det den en enorm fördel i att kunna växa sig till en tumör i kroppen.

I Figur 4.5b observeras att en ökning av reproduktionstakt medför en ökning av fixeringssannolikheten. Detta är väntat eftersom en A-individ med högre reproduktionstakt kan reproducera sig snabbare än en B-individ och på så sätt växer populationen av A-individer till sig snabbare och kan då ta över populationen. Detta framgår i Avsnitt 2.3.1, där det beskrivs att sannolikheten för celledelning med $p_i = r_i \Delta t$. En cell med snabbare reproduktionstakt får därmed en större sannolikhet i att celledelning sker. Resultatet stöds även av resultatet från Moranprocessmodellerna som också visar att en ökad reproduktionstakt leder till en ökad fixeringssannolikhet. Detta innebär att det stämmer för en population med både konstant och varierande storlek.

Figur D.1 och Figur 4.5 visar fixeringssannolikheterna för reproduktionstakt och bärformåga med olika antal gitterpunkter, $L = 10$ och $L = 100$. De är snarlika men det finns en fixeringssannolikhet som skiljer dem åt och det är den första punkten. För $L = 10$, se Figur D.1, har första punkter fixeringssannolikhet på 0.1, medan första punkten för $L = 100$, se Figur 4.5, har en fixeringssannolikhet på 0.01. Resten av punkterna ligger på högre men liknande sannolikheter. Detta beror på att när r_A respektive k_A sätts till 1 så är genotyp A likställd med genotyp B och har alltså inget övertag. Sannolikheterna blir således $1/L$. Detta resultat påminner om resultatet från Moranprocessmodellerna som också får en fixeringssannolikhet på $1/N$ när genotyperna inte konkurrerade med varandra.

Vidare för Modell 4 observeras i Figur 4.7 hur fixeringssannolikheten påverkas då bärformågan och reproduktionstakten förändras samtidigt på två extremvärden av specificiteten. För $\gamma = 0$, då bärformågan delas med omgivningen, fås att bärformågan och reproduktionstakten har ungefär lika stor påverkan eftersom figuren illustrerar en diagonal gradient. Dock ska det finnas i åtanke att reproduktionstakten endast går till att vara två gånger bättre för genotyp A än B, medan bärformågan går till att vara fem gånger bättre för genotyp A än B. Således har reproduktionstakten ett större övertag på fixeringen än bärformågan i det fallet. För $\gamma = 1$ behålls bärformågan hos den egna cellen. Detta gav en tydlig skillnad gentemot $\gamma = 0$. En ökad bärformåga ger en större fördel än en ökad reproduktionstakt eftersom fixeringssannolikheten ökar hastigt för $k_A > 1$. Att en cell inte delar med sig av sin bärformåga ger alltså återigen en stor fördel och i jämförelse är reproduktionstaktens påverkan minimal.

De stokastiska simuleringarna för de tre första modellerna, vilka simulerar en Moranprocess som illustreras i resultatavsnittet, visar sig stämma väl överens med sina individuella analytiska lösningar. Detta tyder på att simuleringarna är tillförlitliga med hänsyn till avgränsningarna som gjorts och vad som studerats. Att ta fram en analytisk lösning till Modell 4 skulle därför vara intressant. Denna skulle då rimligtvis kunna beskrivas med hjälp av en ordinär differentialekvation. Detta gjordes dock inte i projektet då det är en mer avancerad och tidskrävande lösning. Utöver att jämföra resultaten med deras analytiska lösningar skulle det även vara intressant att jämföra dem med faktiska data från riktig cancerforskning och se hur verklighetstrogen den fjärde modellen faktiskt är.

De presenterade modellerna ger således insikt till innebörden av selektion i form av reproduktionstakt och bärformåga för fixering i en population med användning av stokastiska modeller. Då fall av fixering i en population är av intresse ur flera biologiska aspekter, såsom när en mutation uppkommer i en population, är därmed sannolikheten för fixering av en sådan mutant betydelsefull kunskap.

5.2 Diskussion av metodval och avgränsningar

Metoden för modellerna med konstant population baserades på Moranprocessen. I resultatet framgår det att denna metod gav väl överensstämmande resultat med de analytiska lösningarna som togs upp i teorin. Moranprocessen var därmed ett bra sätt att simulera enkla evolutionära processer när endast reproduktionstakt tas i beaktning. När fler faktorer beaktas behövs dock mer komplexa modeller för att kunna beskriva processen. Då valdes en modell som baserades på bröstcancertillväxt som även beaktar bärformåga. I resultatet för cancertillväxt framgår det tydligt att när fler faktorer tas i beaktning ges en mer komplex tillväxt av tumören. Faktorernas inverkan syns tydligt i resultatet.

Valet av metod tycks alltså vara väl anpassat till projektet med hänsyn till dess avgränsningar. Dock är cancer en väldigt komplex sjukdom som beror på många faktorer samtidigt vilket gör det svårt att ta fram en verklighetstrogen modell. Projektets resultat kan därför inte inge tillräcklig tillförlitlighet för att kunna användas i kliniskt syfte. Däremot kan resultaten skapa förståelse över hur olika faktorer såsom reproduktionstakt och bärformåga spelar in i cancerevolution och hur det påverkar sjukdomsförloppet. Studien visar även att specificiteten har en betydande inverkan på dynamiken för konkurrensen mellan de två genotyperna. Detta kan påverka hur effektivt en behandling är och huruvida den kan ge en varaktig effekt på tumörtillväxten.

5.3 Slutsats

Slutsatsen som kan dras är att reproduktionstakten är en parameter som har en stor inverkan på fixeringssannolikheten oberoende typen av selektion. Det framgick även att bärformågan i modelleringen av cancertillväxt har en väsentlig inverkan på fixeringssannolikheten.

Slutmålet i projektet var att skapa en modell som beskriver hur cancerceller kan växa och ta över en population av friska celler. Detta uppnåddes till viss del med beaktning av reproduktionstakt och bärformåga, men då cancer är mer komplext än så är tillförlitligheten till modellen låg. För att få en mer verklighetstrogen modell skulle vidare studier behöva göras där färre avgränsningar beaktas. Det skulle exempelvis vara intressant att studera cancerceller då de konkurrerar om resurser i en tumör, eller så kan rummet beaktas som cellerna befinner sig i för att studera hur det kan se ut när en tumör växer. Det skulle även vara intressant att undersöka uppkomsten av olika mutationer, såsom att bärformågan eller reproduktionstakten kan vara olika stark hos olika celler. Ett annat studieområde av intresse för förståelsen av cancerevolution vore modeller som kan simulera cancerbehandlingar och dess påverkan på cancercellerna i en tumör. Detta då cancerbehandlingar är en drivande faktor i evolutionen av cancer och utvecklandet av cancercellers resistens mot behandlingar. Studier av olika behandlingars påverkan på cancerevolution är således viktiga för framtagning av behandlingar som medför störst sannolikhet att vara framgångsrika och därmed eliminera eller hålla cancer under kontroll. På så sätt hade modellerna blivit mer representativa för cancertillväxt och användbara för vidare forskning.

Referenser

- [1] D. Neill, “Evolution of lifespan”, *Journal of Theoretical Biology*, årg. 358, s. 232–245, okt. 2014. DOI: 10.1016/j.jtbi.2014.06.014.
- [2] G. Folguera och O. Lombardi, “The Relationship between Microevolution and Macroevolution, and the Structure of the Extended Synthesis”, *History and philosophy of the life sciences*, årg. 34, s. 539–559, jan. 2012.
- [3] A. P. Hendry och M. T. Kinnison, “An introduction to microevolution: Rate, pattern, process”, i *Microevolution Rate, Pattern, Process*. Dordrecht, Nederländerna: Springer Netherlands, 2001, s. 1–8.
- [4] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge, MA, USA och London, England: Harvard University Press, 2006.
- [5] C. Offord, “Cancer Researchers Use Evolution to Target Drug Resistance”, *The Scientist*, april 2020.
- [6] P. Gerlee och A. R. Anderson, “The evolution of carrying capacity in constrained and expanding tumour cell populations”, *Physical Biology*, årg. 12, nr 5, juli 2015. DOI: 10.1088/1478-3975/12/5/056001.
- [7] C. Viallard och B. Larrivé, “Tumor angiogenesis and vascular normalization: alternative therapeutic targets”, *Angiogenesis*, årg. 20, nr 4, s. 409–426, 2017. DOI: 10.1007/s10456-017-9562-9.
- [8] F. Wikström, L. Råde och B. Westergren, *Mathematics Handbook: for Science and Engineering*, 6. utg. Lund, Sverige: Studentlitteratur, 2019.

A Appendix - Notationslista

Nedan följer sammanfattande förklaringar av de notationer som används i rapporten.

x	Frekvens av genotyp A i populationen.
v	Förmånen vid vinst av en strid i Hawk- and Dove-spelet.
c	Kostnaden att förlora vid strid i Hawk- and Dove-spelet.
$\pi_{i,j}$	Förväntad payoff för ett möte av två individer i och j i spelteori.
x^*	Jämviktsläge. Beskriver var på intervallet mellan två instabila jämviktslägen som det stabila jämviktsläget hamnar.
$p_{i,j}$	Sannolikheten att gå från ett tillstånd i till ett tillstånd j .
P	Övergångsmatris. Beskriver sannolikheter att gå från ett tillstånd i till ett tillstånd j .
i	Tillstånd i Moranprocessen som beskriver antalet A-individer i populationen.
N	Totalpopulation.
r	Reproduktionstakt. Hur snabbt förökning av individen sker.
s	Selektionskonstant. Beskriver övertaget som en A-individ har i reproduktionstakt.
F_i^A	Förväntad payoff för genotyp A.
F_i^B	Förväntad payoff för genotyp B.
r_i^A	Reproduktionstakt för genotyp A.
r_i^B	Reproduktionstakt för genotyp B.
w	Selektionsintensitet. Hur stark påverkan selektion har i processen.
L	Antalet gitterpunkter (platser) där celler kan befinna sig för cancertillväxt.
γ	Specificitet. Beskriver hur mycket cellen delar med sig av sin bärformåga till angränsande celler.
$K_{i,l}$	Effektiva bärformågan för cell i på plats l .
k_i	Inre bärformågan för cell i . Antar antingen A-individens eller B-individens värde.
\bar{K}_l	Medelvärdet på inre bärformågan för alla celler på plats l .
N_l	Antalet celler på plats l .
ρ_i^r	Sannolikheten för att cell i utför celledelning.
r_i	Reproduktionstakten för cell i . Antar antingen A-individens eller B-individens värde.
Δt	Tidssteg. I varje tidssteg görs ett nytt försök till delning för varje cell.
$\rho_{i,l}$	Sannolikheten att dottercellen till cell i placerar sig på en viss plats l .
ρ_i^a	Sannolikheten att cell i dör.
δ	Avdödningstakt. Med vilken hastighet celler dör.
P_{a_i}	Absorptionssannolikhet. Sannolikheten att processen fastnar i ett absorberande läge givet ett visst initialtillstånd i .
P_f	Fixeringssannolikhet. Sannolikheten att processen fastnar i ett absorberande tillstånd från endast en unik individ i en starpopulation med individer av annan genotyp.
N_0	Antal individer med genotyp A i populationen vid start.

B Appendix - Definitioner och härledningar

I följande avsnitt finns fördjupad teori i form av definitioner och härledningar.

B.1 Markovkedjor

En Markovkedja är en stokastisk process utan minne, vilket betyder att framtida utfall endast beror på nuvarande tillstånd och inte dåtiden. Detta kallas för Markovegenskapen [8]. En Markovkedja kan med matematiska termer definieras som följande.

Definition A1 (*tids-diskret Markovprocess*)

En diskret Markovkedja $\{X_n, n \geq 0\}$ är en följd slumpvariabler X_0, X_1, \dots som har egenskapen $P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i) = P(X_{n+1} = j | X_n = i)$, där i, j är tillstånd processen kan befinna sig i [8].

Definition A2 (*Födelse-dödsprocess*)

En födelse-döds process är en en-dimensionell stokastisk process på ett diskret tillståndsrum $i = 0, \dots, N$. I varje slumpmässig händelse så kan tillståndsvariabeln i förbli samma, röra sig till $i + 1$ eller $i - 1$. Denna process har Markovegenskapen enligt Definition A1 och är tidsdiskret [4].

B.2 Konfidensintervall

Ett konfidensintervall är ett intervall konstruerat med samplade data av en bestämd storlek från en population med en särskild sannolikhetsfördelning [8]. Intervallet är konstruerat så att den innehåller den parameter man är intresserad av med en förbestämd sannolikhet. Ett 95 procentigt konfidensintervall betyder att det sanna parametervärdet är inom intervallet med 95 procents säkerhet [8].

För data som är dikotom, det vill säga kan anta två olika värden, beskrivs ett slumpmässigt utfall med en Bernoullifördelning med parameter $p \in [0, 1]$ [8]. För Bernoullifördelningen fås dess väntevärde och varians givet av $\mu = p$ respektive $\sigma^2 = p(1 - p)$. Det skattade väntevärdet \hat{p} är givet av

$$\hat{p} = \frac{p_1 + \dots + p_n}{n} \quad (\text{B.1})$$

där (p_1, \dots, p_n) är ett stickprov av utfall och n är stickprovets storlek [8]. Det skattade standardfelet är givet av

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n - 1}}. \quad (\text{B.2})$$

Ett approximativt $100(1 - \alpha)\%$ tväsidigt konfidensintervall för p [8] är givet av

$$I_p \approx \hat{p} \pm z_{\alpha/2} \cdot s_{\hat{p}}. \quad (\text{B.3})$$

B.3 Härledning av generell formel för absorptionssannolikhet

Nedan följer en härledning av den generella formeln för absorptionssannolikhet, P_{a_i} , beskriven i Avsnitt 2.4. Absorptionssannolikheten P_{a_i} beskriver sannolikheten att nå tillstånd N från starttillstånd i [4]. Eftersom $i = 0$ och $i = N$ är absorberande tillstånd blir absorptionssannolikheten utifrån dessa tillstånd följande

$$P_{a_0} = 0, \quad P_{a_N} = 1. \quad (\text{B.4})$$

Alla övriga absorptionssannolikheter kan beskrivas enligt

$$P_{a_i} = p_{i,i-1}P_{a_{i-1}} + (1 - p_{i,i+1} - p_{i,i+1})P_{a_i} + p_{i,i+1}P_{a_{i+1}}, \quad i = 1, \dots, N - 1. \quad (\text{B.5})$$

Omskrivning av Ekvation B.5 ger

$$P_{a_i} = p_{i,i-1}P_{a_{i-1}} + P_{a_i} - p_{i,i+1}P_{a_i} - p_{i,i+1}P_{a_i} + p_{i,i+1}P_{a_{i+1}} \quad (\text{B.6})$$

som vidare kan förenklas enligt

$$p_{i,i-1}(P_{a_{i+1}} - P_{a_i}) = p_{i,i-1}(P_{a_i} - P_{a_{i-1}}) \leftrightarrow P_{a_{i+1}} - P_{a_i} = \frac{p_{i,i-1}}{p_{i,i-1}}(P_{a_i} - P_{a_{i-1}}). \quad (\text{B.7})$$

För att kunna hitta ett uttryck för P_{a_i} definieras först variabeln y_i enligt

$$y_i = P_{a_i} - P_{a_{i-1}}. \quad (\text{B.8})$$

Notera att summering av y_i ger

$$\sum_{i=1}^N y_i = P_{a_1} - P_{a_0} + P_{a_2} - P_{a_1} + \dots + P_{a_N} - P_{a_{N-1}} = P_{a_N} - P_{a_0} = 1. \quad (\text{B.9})$$

θ_i kan definieras enligt

$$\theta_i = \frac{p_{i,i-1}}{p_{i,i+1}}. \quad (\text{B.10})$$

Insättning av y_i och θ_i i Ekvation B.5 ger

$$y_{i+1} = \theta_i y_i \quad (\text{B.11})$$

därmed fås följande uttryck

$$y_1 = P_{a_1}, \quad y_2 = \theta_1 P_{a_1}, \quad y_3 = \theta_1 \theta_2 P_{a_1}, \dots \quad (\text{B.12})$$

y_i kan summeras utifrån Ekvation B.11

$$\sum_{i=1}^N y_i = P_{a_1} + \theta_1 P_{a_1} + \theta_1 \theta_2 P_{a_1} + \dots + \theta_1 \cdot \dots \cdot \theta_{N-1} P_{a_1} \quad (\text{B.13})$$

som kan förenklas till

$$1 = P_{a_1} \left(1 + \theta_1 + \theta_1 \theta_2 + \dots + \prod_{k=1}^{N-1} \theta_k \right). \quad (\text{B.14})$$

Omskrivning av uttrycket ger

$$P_{a_1} = 1 / \left(1 + \sum_{j=1}^{N-1} \prod_{k=1}^j \theta_k \right). \quad (\text{B.15})$$

Från Ekvation B.8 kan absorptionssannolikheten uttryckas enligt

$$P_{a_i} = y_i + P_{a_{i-1}} = y_i + y_{i-1} + P_{a_{i-2}} = y_i + y_{i-1} + \dots + y_1 + P_{a_1} \quad (\text{B.16})$$

där alla y -termer innehåller termen P_{a_1} . Därmed kan uttrycket skrivas om på följande sätt

$$P_{a_i} = P_{a_1} \left(1 + \sum_{j=1}^{i-1} \prod_{k=1}^j \theta_k \right). \quad (\text{B.17})$$

Med Ekvation B.15 kan ekvationen för absorptionssannolikheten uttryckas som

$$P_{a_i} = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^j \theta_k}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j \theta_k} = \frac{1 + \sum_{j=1}^{i-1} \prod_{k=1}^j p_{k,k-1}/p_{k,k+1}}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^j p_{k,k-1}/p_{k,k+1}} \quad \square. \quad (\text{B.18})$$

C Appendix - Källkod

Det här avsnittet innehåller all källkod som framtagits för att generera det erhållna resultatet i rapporten.

C.1 Modell 1 - Moranprocess utan selektion

```
import numpy as np
import matplotlib.pyplot as plt

def steg(i, N):
    """
    Simulerar ett steg i markovkedjan.

    Parametrar:

    i: antal A i populationen
    N: total population

    Returnerar: skillnaden i As population
    """

    # sannolikheten att gå från i till i-1 A
    P_minskar = i*(N-i)/N**2
    # sannolikheten att gå från i till i+1 A
    P_ökar = (N-i)*i/N**2
    # sannolikheten att gå från i till i A
    P_stannar = 1 - P_minskar - P_ökar

    # slumpa fram resultatet i detta steg
    return np.random.choice(a=[-1, 0, 1], p=[P_minskar, P_stannar, P_ökar])

def markov(N_0, N):
    """
    Simulerar en markovkedja. Avslutar när den når ett absorberande tillstånd.

    Parametrar:

    N_0: begynnelsevärde för antalet A
    N: total population

    Returnerar:
    1 om A tar över hela populationen,
    0 om B tar över hela populationen
    """

    # antalet A-individer i är till en början lika med N_0
    i = N_0

    # upprepa tills vi når ett absorberande tillstånd
    while True:
        # om antalet A är lika med totalpopulationen har A
        # tagit över hela populationen
        if i == N:
```

```

        return 1
    # om antalet A är 0 har B tagit över hela populationen
    elif i == 0:
        return 0

    # om vi inte nått ett absorberande tillstånd
    # tar vi ett till steg i markovkedjan
    i += steg(i, N)

def simulera(simuleringar, N_0, N):
    """
    Simulerar markovkedjan ett upprepat antal gånger och returnerar resultaten.

    Parametrar:
    simuleringar: antalet simuleringar som ska göras
    N_0: begynnelsevärde för antalet A
    N: total population

    Returnerar: uppskattning av absorptionssannolikheten
    """

    # till en början har A vunnit 0 gånger
    aVinner = 0
    for i in range(simuleringar):
        # simulera en markovkedja och lägg till 1 till aVinner om
        # A tog över populationen
        aVinner += markov(N_0, N)

    return aVinner/simuleringar

def konfidensintervall(simuleringar, N_0, N):
    """
    Simulerar markovkedjor för olika startpopulationer,
    räknar ut konfidensintervall och skapar LaTeX-tabeller
    för absorptionssannolikheterna.

    Parametrar:
    simuleringar: antalet simuleringar som ska göras för varje startpopulation
    N_0: lista av startpopulationer markovkedjan ska simuleras med
    N: total population

    Returnerar:
    lista med medelvärden,
    lista med beloppen på konfidensintervallen
    """

    # Början på kod för en LaTeX-tabell av konfidensintervallen
    tabell = '\\begin{subtable}[h]{0.5\\textwidth}_\\n_\\centering_\\n'
    tabell += '\\caption{$N=$ + str(N) + '$}_\\n_\\begin{tabular}{c|c}_\\n'
    tabell += '\\toprule_\\n_$N_0$&$_I_{P_a}$_\\n_\\bottomrule_\\n'

```

```

# skapar en tom lista att spara medelvärdena i
medelvärde = np.array([])
# skapar en tom lista att spara beloppen av konfidensintervallen i
intervall = np.array([])

# för varje startvärde i N_0
for i in range(len(N_0)):
    start = N_0[i]
    # simulera markovkedjan 'simuleringar' gånger
    p = simulera(simuleringar, start, N)
    # spara medelvärdet i listan
    medelvärde = np.append(medelvärde, p)
    # beräkna det 95%-igt konfidensintervallet
    sp = 1.96*np.sqrt(p*(1-p)/(simuleringar-1))
    # spara intervallens längd i listan
    intervall = np.append(intervall, sp)

# spara medelvärden och konfidensintervall i tabellen
for i in range(len(N_0)):
    tabell += (i+1)%2*'\\rowcolor{black!20}' + str(N_0[i]) + '&$'
    tabell += str(medelvärde[i]) + '\\pm' + str(intervall[i]) + '$\\\\\\n'

# slutet på tabellen med konfidensintervall
tabell += '\\bottomrule\\n\\end{tabular}\\n\\label{}\\n\\end{subtable}\\n'
# printa tabellen med konfidensintervallen
print('Konfidensintervall\\n', tabell)

return medelvärde, intervall

```

```

def graf(simuleringar, N_0, N):

```

```

    """

```

```

    Skapar figur av simulerade resultat tillsammans med analytisk lösning.

```

```

    Parametrar:

```

```

    simuleringar: antalet simuleringar som ska göras för varje startpopulation

```

```

    N_0: lista av startpopulationer markovkedjan ska simuleras med

```

```

    N: total population

```

```

    """

```

```

# simulera data och räkna ut medelvärden och konfidensintervall för

```

```

# de olika N_0-värdena

```

```

medelvärde, intervall = konfidensintervall(simuleringar, N_0, N)

```

```

# beräkna analytiska lösningar för de olika N_0-värdena

```

```

yReferens = [x/N for x in N_0]

```

```

# plotta den analytiska lösningen

```

```

plt.plot(N_0, yReferens, '#f4a836', label="Analytisk lösning")

```

```

# plotta konfidensintervall för väntevärdesskattningarna

```

```

plt.errorbar(N_0, medelvärde, yerr=intervall, fmt='_', ecolor='b',
             capsize=5, label="Konfidensintervall")

```

```

# plotta skattning av väntevärdena

```



```
plt.plot(N_0, medelvärde, 'ro', label = 'Skattat_väntevärde')

# sätt namn på x-axeln
plt.xlabel('Startvärde $N_0$', fontsize=12)
# sätt namn på y-axeln
plt.ylabel('Absorptionssannolikhet_$P_a$', fontsize=12)
# bestäm x-axelns utseende
plt.xticks(N_0, [str(x) for x in N_0], fontsize=12)
# bestäm y-axelns utseende
plt.yticks(np.arange(0, 1.1, 0.1), fontsize=12)
# visar label på funktionerna
plt.legend(loc='upper_left', fontsize=10)
# visar figur
plt.show()
```

C.2 Modell 2 - Moranprocess med konstant selektion

```
import numpy as np
import matplotlib.pyplot as plt

def steg(i, N, s):
    """
    Simulerar ett steg i markovkedjan.

    Parametrar:

    i:  antal A i populationen
    N:  total population
    s:  selektionskonstant

    Returnerar: skillnaden i As population
    """

    # beräkna reproduktionstakt för A
    r = 1 + s

    # sannolikheten att gå från i till i-1 A
    P_minskar = (i/N)*((N-i)/(r*i + (N-i)))
    # sannolikheten att gå från i till i-1 A
    P_ökar = ((N-i)/N)*(r*i/(r*i + (N-i)))
    # sannolikheten att gå från i till i A
    P_stannar = 1 - P_minskar - P_ökar

    # slumpa fram resultatet i detta steg
    return np.random.choice(a=[-1, 0, 1], p=[P_minskar, P_stannar, P_ökar])

def markov(N_0, N, s):
    """
    Simulerar en markovkedja. Avslutar när den når ett absorberande tillstånd.

    Parametrar:

    N_0:  begynnelsevärde för antalet A
    N:    total population
    s:    selektionskonstant

    Returnerar:
    1 om A tar över hela populationen,
    0 om B tar över hela populationen
    """
    i = N_0

    # upprepa tills vi når ett absorberande tillstånd
    while True:
        # om antalet A är lika med totalpopulationen har A tagit
        # över hela populationen
        if i == N:
            return 1
        # om antalet A är 0 har B tagit över hela populationen
```

```

    elif i == 0:
        return 0

    # om vi inte nått ett absorberande tillstånd tar vi
    # ett till steg i markovkedjan
    i += steg(i, N, s)

def simulera(simuleringar, N_0, N, s):
    """
    Simulerar markovkedjan ett upprepat antal gånger och returnerar resultaten.

    Parametrar:

    simuleringar: antalet simuleringar som ska göras
    N_0: begynnelsevärde för antalet A
    N: total population
    s: selektionskonstant

    Returnerar: uppskattning av absorptionssannolikheten
    """

    # till en början har A vunnit 0 gånger
    aVinner = 0
    for i in range(simuleringar):
        # simulera en markovkedja och lägg till 1 till aVinner om
        # A tog över populationen
        aVinner += markov(N_0, N, s)

    return aVinner/simuleringar

def konfidensintervall(simuleringar, N_0, N, s):
    """
    Simulerar markovkedjor för olika startpopulationer,
    räknar ut konfidensintervall och LaTeX-tabeller
    för absorptionssannolikheterna.

    Parametrar:
    simuleringar: antalet simuleringar som ska göras för varje startpopulation
    N_0: lista av startpopulationer markovkedjan ska simuleras med
    N: total population

    Returnerar: lista med medelvärden, lista med beloppen på konfidensintervallen
    """

    # Början på kod för en LaTeX-tabell av konfidensintervallen
    tabell = '\\begin{subtable}[h]{0.5\\textwidth} \\n \\centering \\n '
    tabell += '\\caption{N=' + str(N) + ', $\\(N_{0})=' + str(N_0) + '\\)} \\n '
    tabell += '\\begin{tabular}{c|c} \\n \\toprule \\n $$ & $I_{P_a}$ \\n \\n '
    tabell += '\\bottomrule \\n '

    # skapar en tom lista att spara medelvärdena i
    medelvärde = np.array([])
    # skapar en tom lista att spara beloppen av konfidensintervallen i

```

```

intervall = np.array ([])

# för varje selektionsvärde i s
for i in range(len(s)):
    selektion = s[i]
    # simulera markovkedjan 'simuleringar' gånger
    p = simulera(simuleringar, N_0, N, selektion)
    # spara medelvärdet i listan
    medelvärde = np.append(medelvärde, p)
    # beräkna det 95%-igt konfidensintervallet
    sp = 1.96*np.sqrt(p*(1-p)/(simuleringar-1))
    # spara intervallets längd i listan
    intervall = np.append(intervall, sp)

# spara medelvärden och konfidensintervall i tabellen
for i in range(len(s)):
    tabell += (i+1)%2*'\\rowcolor{black!20}' + str(s[i]) + ' & $'
    tabell += str(medelvärde[i]) + ' \pm ' + str(intervall[i]) + '$ \\ \\ \\n'

# slutet på tabellen med konfidensintervall
tabell += '\\bottomrule \\n \\end{tabular} \\n \\label{} \\n \\end{subtable} \\n'
# printa tabellen med konfidensintervallen
print('Konfidensintervall \\n', tabell)

return medelvärde, intervall

def graf(simuleringar, N_0, N, s):
    """
    Skapar figur av simulerade resultat tillsammans med analytisk lösning.

    Parametrar:
    simuleringar: antalet simuleringar som ska göras för varje startpopulation
    N_0: lista av startpopulationer markovkedjan ska simuleras med
    N: total population
    s: selektionskonstant

    """

    # simulera data och räkna ut medelvärden och konfidensintervall för
    # de olika s-värdena
    medelvärde, intervall = konfidensintervall(simuleringar, N_0, N, s)
    # beräkna analytiska lösningar för de olika s-värdena
    yReferens = [1/N if x==0 else (1-(1/(1+x))) / (1-(1/(1+x)**N))] for x in s]

    # plotta den analytiska lösningen
    plt.plot(s, yReferens, '#f4a836', label="Analytisk lösning")
    # plotta 1/N
    plt.plot(s, [1/N]*len(s), 'c:', label="$1/N$")
    # plotta konfidensintervall för väntevärdesskattningarna
    plt.errorbar(s, medelvärde, yerr=intervall, fmt=' ', ecolor='b',
    capsize=5, label="Konfidensintervall")
    # plotta skattning av väntevärdena
    plt.plot(s, medelvärde, 'ro', label='Skattat väntevärde')

```

```
# sätt namn på x-axeln
plt.xlabel('Selektionskonstant  $s$ ', fontsize=12)
# sätt namn på y-axeln
plt.ylabel('Fixeringssannolikhet  $P_f$ ', fontsize=12)
# bestäm x-axelns utseende
plt.xticks(s, [str(x) for x in s], fontsize=12)
# bestäm y-axelns utseende
plt.yticks(np.arange(0, 1.1, 0.1), fontsize=12)
# visar label på funktionerna
plt.legend(loc='upper left', fontsize=10)
# visar figur
plt.show()
```

C.3 Modell 3 - Moranprocess med frekvensberoende selektion

```
import numpy as np
import matplotlib.pyplot as plt

def kategorisera(w, v, c):
    """
    Avgör vilken slags jämförelse som ska göras.

    Parametrar:
    w:   selektionsintensitet
    v:   vinst för nytt territorie
    c:   kostnad vid skada

    Returnerar:
    namn:      namnet på variabeln som varierar
    fulltNamn: fulla namnet på variabeln som varierar
    varierar:  variabeln som varierar
    """

    # kolla vilka av parametrarna har längd>1
    indikator = [True if len(x)>1 else False for x in [w, v, c]]

    # avbryt med felmeddelande om inte exakt en parameter har >1 värden
    if sum(indikator)!=1:
        raise Exception("Variera_en_variabel_i_taget")

    # hitta index på den parameter som har >1 värden
    i = indikator.index(True)
    # bestäm variabelns namn
    namn = ["w", "v", "c"][i]
    # bestäm variabelns fulla namn
    fulltNamn = ["Selektionsintensitet_$$", "Vinstvärde $$",
    "Kostnadsvärde $$"][i]
    # spara variabeln
    varierar = [w, v, c][i]

    return namn, fulltNamn, varierar

def reproduktionstakt(i, N, w, v, c):
    """
    Beräknar den relativa reproduktionstakten.

    Parametrar:
    i:   antalet hökar
    N:   total population
    w:   selektionsintensitet
    v:   vinst för nytt territorie
    c:   kostnad vid skada

    Returnerar: den relativa reproduktionstakten  $r = rA_i/rB_i$ 
    """

    # förväntad pay-off för hökarna
    FA_i = ((i - 1)*(v/2 - c/2) + (N - i)*v) / (N - 1)
    # förväntad pay-off för duvorna
```

```

FB_i = (N - i - 1)*(v/2) / (N - 1)

# reproduktionstakt för hökarna
rA_i = 1 - w + w*FA_i
# reproduktionstakt för duvorna
rB_i = 1 - w + w*FB_i

# returnera relativ reproduktionstakten
return rA_i/rB_i

def steg(i, N, w, v, c):
    """
    Simulerar ett steg i markovkedjan.

    Parametrar:

    i: antal hökar i populationen
    N: total population
    w: selektionsintensitet
    v: vinst för nytt territorie
    c: kostnad vid skada

    Returnerar: skillnaden i hökarnas population
    """

    # beräkna den relativa reproduktionstakten
    r = reproduktionstakt(i, N, w, v, c)

    # sannolikheten att gå från i till i-1 hökar
    P_minskar = (N-i)*i/(N*(i*r + N - i))
    # sannolikheten att gå från i till i-1 hökar
    P_ökar = i*(N-i)*r/(N*(i*r + N - i))
    # sannolikheten att gå från i till i hökar
    P_stannar = 1 - P_minskar - P_ökar

    # slumpa fram ett utfall utifrån sannolikheterna ovan och returnera det
    return np.random.choice(a=[-1, 0, 1], p=[P_minskar, P_stannar, P_ökar])

def markov(N, w, v, c):
    """
    Simulerar en markovkedja. Avslutar när den når ett absorberande tillstånd.

    Parametrar:

    N: total population
    w: selektionsintensitet
    v: vinst för nytt territorie
    c: kostnad vid skada

    Returnerar:
    1 om hökarna tar över hela populationen,
    0 om duvorna tar över hela populationen
    """
    # antalet hökar i är till en början 1

```

```

i = 1

# upprepa tills vi når ett absorberande tillstånd
while True:
    # om antalet hökar är lika med totalpopulationen har
    # hökarna tagit över hela populationen
    if i == N:
        return 1
    # om antalet hökar är 0 har duvorna tagit över hela populationen
    elif i == 0:
        return 0

    # om vi inte nått ett absorberande tillstånd tar vi ett
    # till steg i markovkedjan
    i += steg(i, N, w, v, c)

def simulera(simuleringar, N, w, v, c):
    """
    Simulerar markovkedjan ett upprepat antal gånger och returnerar resultaten.

    Parametrar:

    simuleringar: antalet simuleringar som ska göras
    N: total population
    w: selektionsintensitet
    v: vinst för nytt territorie
    c: kostnad vid skada

    Returnerar: uppskattning av absorptionssannolikheten
    """

    # till en början har hökarna vunnit 0 gånger
    hökarVinner = 0
    for i in range(simuleringar):
        # simulera en markovkedja och lägg till 1 till hökarVinner om
        # hökarna tog över populationen
        hökarVinner += markov(N, w, v, c)

    return hökarVinner/simuleringar

def numeriskFixering(N, w, v, c):
    """
    Beräknar fixeringssannolikheten numeriskt.

    Parametrar:

    N: total population
    w: selektionsintensitet
    v: vinst för nytt territorie
    c: kostnad vid skada

    Returnerar: lista med y-värden som korresponderar mot de olika v värdena
    """

    # skapa en tom lista att spara sannolikheterna i

```



```

resultat = []

# beräkna fixeringssannolikheten för varje parameterkombination
for vinst in v:
    for kostnad in c:
        for intensitet in w:
            # se avsnitt 2.4.3 i rapporten för formeln vi använder för
            # att beräkna fixeringssannolikheten
            sum_ki = 0
            for k in range(1, N):
                prod_ki = 1
                for i in range(1, k+1):
                    prod_ki *= 1/reproduktionstakt(i, N, intensitet,
                    vinst, kostnad)
                sum_ki += prod_ki
            result = 1 / (1 + sum_ki)
            # lägg till den beräknade fixeringssannolikheten i listan
            resultat.append(result)

return resultat

def konfidensintervall(simuleringar, N, w, v=[1], c=[1]):
    """
    Simulerar markovkedjor för olika startpopulationer
    och räknar ut konfidensintervall för fixeringssannolikheterna.
    Skapar en LaTeX-tabell med konfidensintervallen.

    Parametrar:
    simuleringar: antalet simuleringar som ska göras för varje startpopulation
    N: total population
    w: selektionsintensitet
    v: vinst för nytt territorie
    c: kostnad vid skada

    Returnerar: lista med medelvärden, lista med beloppen på konfidensintervallen
    """

    # avgör vilken typ av simulation vi ska göra och
    # vilken parameter som ska varieras
    namn, fulltNamn, varieras = kategorisera(w, v, c)

    # Början på kod för en LaTeX-tabell av konfidensintervallen
    tabell = '\\begin{subtable}[h]{0.49\\textwidth}_\\n_\\centering'
    tabell += '\\n_\\caption{\\$w=' + str(w) + '\\$}_\\n_\\begin{tabular}{c|c}_\\n_'
    tabell += '\\n_\\toprule_\\n_\\$' + namn + '\\$&\\$I_{P_f}\\$_\\n_\\bottomrule_\\n_'

    # skapar en tom lista att spara medelvärden i
    medelvärde = np.array([])
    # skapar en tom lista att spara beloppen av konfidensintervallen i
    intervall = np.array([])

    # för alla parameterkombinationer
    for intensitet in w:
        for vinst in v:
            for kostnad in c:

```

```

# simulera markovkedjan 'simuleringar' gånger
p = simulera(simuleringar, N, intensitet, vinst, kostnad)
# spara medelvärdet i listan
medelvärde = np.append(medelvärde, p)
# beräkna det 95%-igt konfidensintervallet
sp = 1.96*np.sqrt(p*(1-p)/(simuleringar-1))
# spara intervallens längd i listan
intervall = np.append(intervall, sp)

# spara medelvärden och konfidensintervall i tabellen
for i in range(len(varierar)):
    tabell += (i+1)%2*'\\rowcolor{black!20}' + str(varierar[i]) + '&$'
    tabell += str(medelvärde[i]) + '\\pm' + str(intervall[i]) + '$\\\\\\n'

# slutet på tabellen med konfidensintervall
tabell += '\\bottomrule\\n\\end{tabular}\\n\\label{}\\n\\end{subtable}\\n'

# printa tabellen med konfidensintervallen
print('Konfidensintervall\\n', tabell)

return medelvärde, intervall, fulltNamn, varierar

def graf(simuleringar, N, w, v, c):
    """
    Skapar figur av simulerade resultat tillsammans med analytisk lösning.

    Parametrar:
    simuleringar: antalet simuleringar som ska göras för varje startpopulation
    N: total population
    w: selektionsintensitet
    v: vinst för nytt territorie
    c: kostnad vid skada

    """

    # beräkna lösningen numeriskt
    yReferens = numeriskFixering(N, w, v, c)
    # simulera data och räkna ut medelvärden och konfidensintervall för
    # de olika s-värdena
    medelvärde, intervall, fulltNamn, varierar = konfidensintervall(simuleringar,
    N, w, v, c)

    # plotta den numeriska lösningen
    plt.plot(varierar, yReferens, '#f4a836', label='Numerisk lösning')
    # plotta 1/N
    plt.plot(varierar, [1/N]*len(varierar), 'c:', label="$1/N$")
    # plotta konfidensintervall för väntevärdesskattningarna
    plt.errorbar(varierar, medelvärde, yerr=intervall, fmt='_', ecolor='b',
    capsiz=5, label="Konfidensintervall")
    # plotta skattning av väntevärdena
    plt.plot(varierar, medelvärde, 'ro', label='Skattat väntevärde')

    # sätt namn på x-axeln
    plt.xlabel(fulltNamn, fontsize=12)

```

```
# sätt namn på y-axeln
plt.ylabel("Fixeringssannolikhet_{$P_f}$", fontsize=12)
# bestäm x-axelns utseende
plt.xticks(varierar, [str(x) for x in varierar], fontsize=12)
# bestäm y-axelns utseende
plt.yticks(np.arange(0, 0.61, 0.1), fontsize=12)
# visar label på funktionerna
plt.legend(loc='upper_left', fontsize=10)
# visar figur
plt.show()
```

C.4 Modell 4 - Modell för concertillväxt

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from numba import njit

def kategorisera(k_A, r_A):
    """
    Avgör vilken slags jämförelse som ska göras.

    Parametrar:
    w - selektionsintensitet
    V - vinst för nytt territorie
    C - kostnad vid skada

    Returnerar:
    namn - namnet på variabeln som varierar
    fulltNamn - fulla namnet på variabeln som varierar
    typ - typen av beräkning som görs
    fullTyp - fulla namnet på typen av beräkning som görs
    varierar - variabeln som varierar
    """

    # kolla vilka av parametrarna har längd>1
    indikator = [True if len(x)>1 else False for x in [k_A, r_A]]

    # avbryt med felmeddelande om inte exakt en parameter har >1 värden
    if sum(indikator)!=1:
        raise Exception("Variera_en_variabel_i_taget")

    # hitta index på den parameter som har >1 värden
    i = indikator.index(True)
    # bestäm variabelns namn
    namn = ['$k_A$', '$r_A$'][i]
    # bestäm variabelns fulla namn
    fulltNamn = ['Bärförmåga $k_A$', 'Reproduktionstakt_$r_A$'][i]
    # spara variabeln
    varierar = [k_A, r_A][i]

@njit
def delning(L, X, r_X, k_X, Y, k_Y, delta, gamma, dt):
    """
    Simulerar ett försök till delning av varje cell i en endimensionella kanal

    Parametrar:

    L: antalet gitterpunkter
    X: lista med population av typ X
    r_X: reproduktionstakt för X
    k_X: bärförmåga för X
    Y: lista med population av typ Y
    k_Y: bärförmåga för Y
    delta: dödlighet
    """
```

```

gamma: specificitet
dt: tidssteg

Returnerar: lista med förändring i population av typ X
'''

# skapar en tom lista för att lagra förändringar
förändring = np.zeros(L)

# loopar igenom alla gitterpunkter
for l in range(L):

    # loopar igenom alla celler på plats l
    for i in range(int(X[l])):

        # slumpa fram om cellen ska delas eller inte
        dela = np.random.random()

        # om cellen ska delas
        if dela < r_X*dt:

            # beräkna antalet celler på plats l
            N_l = X[l] + Y[l]
            # beräkna medelkapaciteten för cellerna på plats l
            K_l = (1/N_l)*(k_X*X[l] + k_Y*Y[l])
            # beräkna den specifika kapaciteten för cell i på plats l
            Ki_l = gamma*k_X + (1-gamma)*K_l
            # sannolikheten att dottercellen stannar
            piStanna = 1 - N_l/Ki_l

            # om sannolikheten är negativ, sätt den till 0
            if piStanna < 0:
                piStanna = 0

            # slumpa fram om dottercellen ska stanna eller inte
            slump = np.random.binomial(n=1, p=piStanna)

            # om dottercellen stannar
            if slump == 1:
                # uppdatera förändringen
                förändring[l] += 1
            # om dottercellen inte stannar
            else:
                # genererar start-end connection
                a = np.where(l == L-1, 0, l+1)

                # beräkna antalet celler på plats l+1
                N_lFramåt = X[a] + Y[a]
                # beräkna antalet celler på plats l-1
                N_lBakåt = X[l-1] + Y[l-1]

                # beräkna medelkapaciteten för cellerna på plats l+1
                if N_lFramåt == 0:
                    K_lFramåt = k_X
                else:
                    K_lFramåt = (1/N_lFramåt)*(k_X*X[a] + k_Y*Y[a])

```

```

# beräkna medelkapaciteten för cellerna på plats l-1
if N_lBakåt == 0:
    K_lBakåt = k_X
else:
    K_lBakåt = (1/N_lBakåt)*(k_X*X[l-1] + k_Y*Y[l-1])

# beräkna den specifika kapaciteten för cell i på plats l+1
Ki_lFramåt = gamma*k_X + (1-gamma)*K_lFramåt
# beräkna den specifika kapaciteten för cell i på plats l-1
Ki_lBakåt = gamma*k_X + (1-gamma)*K_lBakåt

# sannolikheten att dottercellen flyttar framåt
piFramåt = 1 - N_lFramåt/Ki_lFramåt

# om sannolikheten är negativ, sätt den till 0
if piFramåt < 0:
    piFramåt = 0

# sannolikheten att dottercellen flyttar bakåt
piBakåt = 1 - N_lBakåt/Ki_lBakåt

# om sannolikheten är negativ, sätt den till 0
if piBakåt < 0:
    piBakåt = 0

# slumpa fram om dottercellen ska testa att flytta framåt
# eller bakåt först
slump = np.random.binomial(n=1, p=0.5)

# om dottercellen testat att flyttar bakåt
if slump == 1:
    # slumpa fram om dottercellen flyttar bakåt eller inte
    slump = np.random.binomial(n=1, p=piBakåt)
    # om dottercellen flyttar bakåt
    if slump == 1:
        # uppdatera förändringen
        förändring[l-1] += 1
    # om dottercellen inte flyttar bakåt
    else:
        # slumpa fram om dottercellen flyttar framåt eller inte
        slump = np.random.binomial(n=1, p=piFramåt)
        # om dottercellen flyttar framåt
        if slump == 1:
            # uppdatera förändringen
            förändring[a] += 1
# om dottercellen testat att flyttar framåt
else:
    # slumpa fram om dottercellen flyttar framåt eller inte
    slump = np.random.binomial(n=1, p=piFramåt)
    # om dottercellen flyttar framåt
    if slump == 1:
        # uppdatera förändringen
        förändring[a] += 1
    # om dottercellen inte flyttar framåt
    else:

```

```

        # slumpa fram om dottercellen flyttar bakåt eller inte
        slump = np.random.binomial(n=1, p=piBakåt)
        # om dottercellen flyttar bakåt
        if slump == 1:
            # uppdatera förändringen
            förändring[1-1] += 1

    # slumpa fram om cellen ska dö
    dö = np.random.random()

    # om cellen ska dö
    if dö < delta*dt:
        # uppdatera förändringen
        förändring[1] -= 1

    return förändring

@njit
def modell(L, r_B, r_A, k_B, k_A, delta, gamma, dt):
    """
    Simulerar en cancermodell. Avslutar när den når ett absorberande tillstånd.

    Parametrar:

    L: antalet gitterpunkter
    r_B: reproduktionstakt för B
    r_A: reproduktionstakt för A
    k_B: bärförmåga för B
    k_A: bärförmåga för A
    delta: dödlighet
    gamma: specificitet
    dt: tidssteg

    Returnerar: 1 om A tar över populationen, 0 om B tar över populationen
    """

    # skapar en listor för B
    B = np.ones(L)
    # skapar en lista för A
    A = np.zeros(L)

    # genererar en slumpmässig mutation i populationen
    slumpad = np.random.randint(0,L)

    # uppdaterar listorna för B och A med mutationen
    B[slumpad] = 0
    A[slumpad] = 1

    # upprepar tills någon genotyp tagit över populationen
    while True:
        if np.all(B==0):
            # om det inte finns kvar några B har A tagit över populationen
            return 1
        elif np.all(A==0):

```

```

        # om det inte finns kvar några A har B tagit över populationen
        return 0

    # om ingen tagit över populationen görs ett nytt försök på delning för B
    B = np.add(delning(L, B, r_B, k_B, A, k_A, delta, gamma, dt), B)
    # om ingen tagit över populationen görs ett nytt försök på delning för A
    A = np.add(delning(L, A, r_A, k_A, B, k_B, delta, gamma, dt), A)

def simulera(simuleringar, L, r_B, r_A, k_B, k_A, delta, gamma, dt):
    """
    Simulerar modellen ett upprepat antal gånger och returnerar resultaten.

    Parametrar:
    simuleringar: antalet simulationer som ska göras
    L: antalet gitterpunkter
    r_B: reproduktionstakt för B
    r_A: reproduktionstakt för A
    k_B: bärförmåga för B
    k_A: bärförmåga för A
    delta: dödlighet
    gamma: specificitet
    dt: tidssteg

    Returnerar: uppskattning av fixeringssannolikheten
    """

    # till en början har A vunnit 0 gånger
    aVinner = 0
    for _ in range(simuleringar):
        # simulera modellen och lägg till 1 till aVinner om A tog
        # över populationen
        aVinner += modell(L, r_B, r_A, k_B, k_A, delta, gamma, dt)

    return aVinner/simuleringar

    return namn, fulltNamn, varierar

def konfidensintervall(simuleringar, L, r_B, r_A, k_B, k_A, delta, gamma, dt):
    """
    Simulerar markovkedjor för olika startpopulationer och räknar ut
    konfidensintervall för fixeringssannolikheterna.
    Skapar en LaTeX-tabell med konfidensintervallen.

    Parametrar:
    simuleringar: antalet simuleringar som ska göras för varje startpopulation.
    L: antalet gitterpunkter
    r_B: reproduktionstakt för B
    r_A: reproduktionstakt för A
    k_B: bärförmåga för B
    k_A: bärförmåga för A
    delta: dödlighet
    gamma: specificitet
    dt: tidssteg
    """

```



```

k_A: bärförmåga för A
delta: dödlighet
gamma: specificitet
dt: tidssteg

"""

# skapa en lista med färger
former = ['o', 's', 'v']
färger = ['r', 'g', 'c']

for (specificitet, form, färg) in zip(gamma, former, färger):
    # simulera data och räkna ut medelvärden och konfidensintervall
    # för de olika s-värdena
    medelvärde, intervall, fulltNamn, varierar = konfidensintervall
    (simuleringar, L, r_B, r_A, k_B, k_A, delta, specificitet, dt)
    # plotta konfidensintervall för väntevärdesskattningarna
    if färg == 'r':
        plt.errorbar(varierar, medelvärde, yerr=intervall, fmt='_',
                     ecolor='b', capsize=5, label="Konfidensintervall")
    else:
        plt.errorbar(varierar, medelvärde, yerr=intervall, fmt='_',
                     ecolor='b', capsize=5)
    # plotta skattning av väntevärdena
    plt.plot(varierar, medelvärde, färg + form,
             label='Skattat_väntevärde, $\gamma$='+str(specificitet))

# sätt namn på x-axeln
plt.xlabel(fulltNamn, fontsize=12)
# sätt namn på y-axeln
plt.ylabel('Fixerings sannolikhet_ $P_f$', fontsize=12)
# bestäm x-axelns utseende
plt.xticks(varierar, [str(x) for x in varierar], fontsize=12)
# bestäm y-axelns utseende
plt.yticks(np.arange(0, 1.1, 0.1), fontsize=12)
# visar label på funktionerna
plt.legend(loc='upper_left', fontsize=12)
# visar figur
plt.show()

```

```

def färgdiagram(simuleringar, L, r_B, r_A, k_B, k_A, delta, gamma, dt):
    """

```

Skapar figur i form av färgdiagram av simulerade resultat.

Parametrar:

simuleringar: antalet simuleringar som ska göras för varje startpopulation

L: antalet gitterpunkter

r_B: reproduktionstakt för B

r_A: reproduktionstakt för A

k_B: bärförmåga för B

k_A: bärförmåga för A

delta: dödlighet

gamma: specificitet

```

dt: tidssteg

"""

# skapa tom data
data = {}

if len(r_A) > 1:
    for reproduktionstakt in r_A:
        # simulera data och räkna ut medelvärden för olika bärförmågor
        medelvärde, _, _, _ = konfidensintervall(simuleringar, L, r_B,
                                                [reproduktionstakt], k_B, k_A,
                                                delta, gamma[0], dt)

        # lägg till i data
        data[reproduktionstakt] = medelvärde
        # bestäm namn
        namn = 'Reproduktionstakt_{}_r_A$'.format(reproduktionstakt)
elif len(gamma) > 1:
    for specificitet in gamma:
        # simulera data och räkna ut medelvärden för olika bärförmågor
        medelvärde, _, _, _ = konfidensintervall(simuleringar, L, r_B,
                                                r_A, k_B, k_A, delta, specificitet, dt)

        # lägg till i data
        data[specificitet] = medelvärde
        # bestäm namn
        namn = 'Specificitet_{}_gamma$'.format(specificitet)
else:
    print("Variera_en_variabel_i_taget")

    # indexering av x-axeln
    index = [str(x) for x in k_A]
    # skapa en dataframe av data
    df = pd.DataFrame(data, index)
    # transponera dataframe
    df_transposed = df.transpose()

    # skapa en färgdiagram
    färgdiagram = sns.färgdiagram(df_transposed, cmap="Reds", linewidth=0,
                                  vmin=0, vmax=1)

    # invertera y-axeln
    färgdiagram.invert_yaxis()
    # sätt namn på colorbar
    färgdiagram.collections[0].colorbar.set_label('$P_f$', rotation=0, ha='left')

    # sätt namn på x-axeln
    plt.xlabel('Bärförmåga $k_A$')
    # sätt namn på y-axeln
    plt.ylabel(namn)
    # rotera y-axelns etiketter
    plt.yticks(rotation=0)
    # visa figur
    plt.show()

```

D Appendix - Rådata

I följande avsnitt redovisas rådata som framtagits för resultatet.

D.1 Modell 1 - Moranprocess utan selektion

Tabell D.1 innehåller skattat väntevärde samt 95 % konfidensintervall för absorptions sannolikhet för Moranprocess utan selektion. Den beskriver intervallen för de olika populationsstorlekarna som undersöktes, vilka var $N = 10$ och $N = 100$, med ett varierande värde på N_0 .

Tabell D.1: Skattat väntevärde samt tvåsidigt 95 % konfidensintervall för absorptions sannolikhet P_a för Modell 1 med varierande startvärde N_0 för de olika populationsstorlekar (a) $N = 10$ och (b) $N = 100$.

(a)		(b)	
N_0	I_{P_a}	N_0	I_{P_a}
1	$0.099 \pm 0.018520526134218982$	10	$0.097 \pm 0.018352831171566202$
2	$0.197 \pm 0.024664039824417665$	20	$0.193 \pm 0.024473086719365284$
3	$0.314 \pm 0.02878062422413264$	30	$0.304 \pm 0.028524283148383046$
4	$0.407 \pm 0.030464768236500725$	40	$0.397 \pm 0.030340814600643643$
5	$0.5 \pm 0.031005827861248298$	50	$0.49 \pm 0.030999626075435438$
6	$0.611 \pm 0.030232129068724682$	60	$0.594 \pm 0.030452963819723778$
7	$0.687 \pm 0.028755694733036435$	70	$0.689 \pm 0.028705368916016473$
8	$0.813 \pm 0.024179059463025367$	80	$0.801 \pm 0.024758032202874255$
9	$0.905 \pm 0.018182743802082568$	90	$0.904 \pm 0.01826809069217518$

D.2 Modell 2 - Moranprocess med konstant selektion

I Tabell D.2 återfinns skattat väntevärde samt 95 % konfidensintervall för fixeringssannolikhet, som erhöles från en Moranprocess med konstant selektion. Tabellen beskriver intervallen för populationsstorlekarna som undersöktes, vilka var $N = 10$ och $N = 100$, samt selektionskonstant på intervallet $s = \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5\}$.

Tabell D.2: Skattat väntevärde samt tvåsidigt 95 % konfidensintervall för fixeringssannolikhet P_f för Modell 2 med varierande selektionskonstant s för olika populationsstorlekar (a) $N = 10$ och (b) $N = 100$.

(a)		(b)	
s	I_{P_f}	s	I_{P_f}
0	$0.106 \pm 0.019089520480939072$	0	$0.009 \pm 0.005856409132559638$
0.5	$0.321 \pm 0.028950810763152104$	0.5	$0.324 \pm 0.02902145487812841$
1	$0.496 \pm 0.031004835658881246$	1	$0.493 \pm 0.03100278914121332$
1.5	$0.588 \pm 0.03052183205234954$	1.5	$0.592 \pm 0.03047644190372477$
2	$0.663 \pm 0.02931197232056078$	2	$0.672 \pm 0.029113531275049807$
2.5	$0.735 \pm 0.027367802919427725$	2.5	$0.723 \pm 0.027751237896800295$
3	$0.716 \pm 0.027963337759621235$	3	$0.757 \pm 0.026596494790387983$
3.5	$0.751 \pm 0.02681593654625646$	3.5	$0.786 \pm 0.02543264252305894$
4	$0.809 \pm 0.024376102739871782$	4	$0.812 \pm 0.02422870858960336$
4.5	$0.831 \pm 0.023238986916664695$	4.5	$0.788 \pm 0.025345704455278693$
5	$0.817 \pm 0.02397783130297638$	5	$0.827 \pm 0.023455739709617462$
5.5	$0.859 \pm 0.021581392608145258$	5.5	$0.839 \pm 0.022791202805585435$

D.3 Modell 3 - Moranprocess med frekvensberoende selektion

Tabell D.3 innehåller skattat väntevärde samt 95 % konfidensintervall för fixeringssannolikhet för Moranprocess med frekvensberoende selektion vid samexistens. De varierande parametrarna var selektionsintensiteten w som antog värdena $w = 0.1$ och $w = 0.5$, och vinstvärdet på intervallet

$v = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$. De fixa parametervärdena var totalpopulationen $N = 100$ samt kostnaden $c = 1$.

Tabell D.3: Skattat väntevärde samt tvåsidigt 95 % konfidensintervall för fixeringssannolikhet P_f för Modell 3 vid samexistens med varierande vinstvärde v för selektionsintensiteterna (a) $w = 0.1$ och (b) $w = 0.5$. De konstanta parametrarna var populationsstorlek $N = 100$ och kostnadsvärde $c = 1$.

(a)		(b)	
v	I_{P_f}	v	I_{P_f}
0.1	$0.004 \pm 0.0039141096886335025$	0.1	0.0 ± 0.0
0.2	$0.004 \pm 0.0039141096886335025$	0.2	0.0 ± 0.0
0.3	$0.004 \pm 0.0039141096886335025$	0.3	0.0 ± 0.0
0.4	$0.012 \pm 0.006752155300435646$	0.4	0.001 ± 0.00196
0.5	$0.012 \pm 0.006752155300435646$	0.5	$0.053 \pm 0.013892687009803121$
0.6	$0.017 \pm 0.008016310799790564$	0.6	$0.176 \pm 0.023615289121251097$
0.7	$0.023 \pm 0.009295751965527297$	0.7	$0.183 \pm 0.023977831302976378$
0.8	$0.038 \pm 0.011856395055146556$	0.8	$0.213 \pm 0.025389286430805128$
0.9	$0.037 \pm 0.011705428939883692$	0.9	$0.237 \pm 0.026369978106652748$

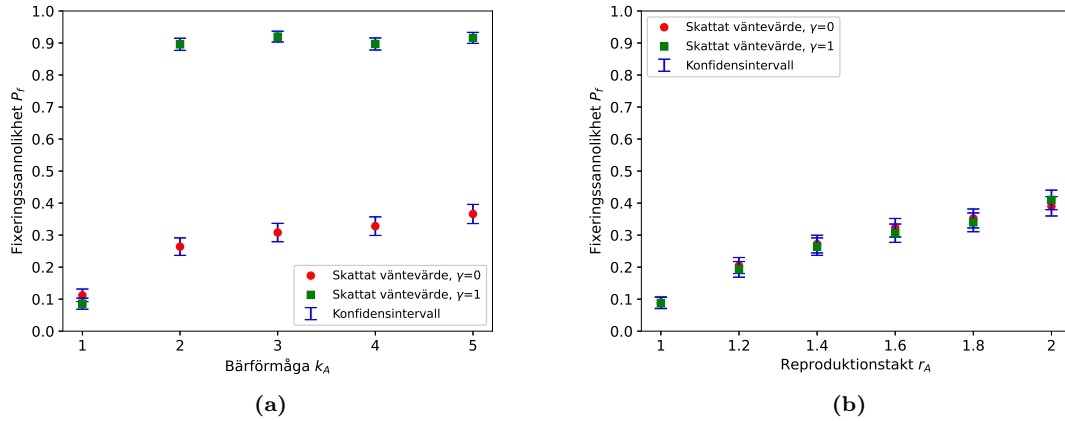
Tabell D.4 visar skattat väntevärde samt 95 % konfidensintervall för fixeringssannolikhet för en Moranprocess med frekvensberoende selektion vid dominans där de varierande parametrarna var selektionsintensiteten w som antog värdena $w = 0.1$ och $w = 0.5$, och vinstvärdet på intervallet $v = \{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$. De fixa parametervärdena var totalpopulationen $N = 100$, samt kostnaden $c = 1$.

Tabell D.4: Skattat väntevärde samt tvåsidigt 95 % konfidensintervall för fixeringssannolikhet P_f för Modell 3 vid dominans med varierande vinstvärde v för selektionsintensiteterna (a) $w = 0.1$ och (b) $w = 0.5$. De konstanta parametrarna var populationsstorlek $N = 100$ och kostnadsvärde $c = 1$.

(a)		(b)	
v	I_{P_f}	v	I_{P_f}
1	$0.036 \pm 0.011552157319684437$	1	$0.254 \pm 0.0269935248677305$
1.5	$0.07 \pm 0.01582208894231411$	1.5	$0.287 \pm 0.028051690625467602$
2	$0.08 \pm 0.016823340476397212$	2	$0.323 \pm 0.028998058572963142$
2.5	$0.098 \pm 0.018436973601857642$	2.5	$0.378 \pm 0.030068683897892027$
3	$0.125 \pm 0.020508427428635175$	3	$0.387 \pm 0.030203623432768268$
3.5	$0.171 \pm 0.023347944254278597$	3.5	$0.398 \pm 0.030353802841603668$
4	$0.161 \pm 0.022791202805585435$	4	$0.426 \pm 0.030664371868702973$
4.5	$0.168 \pm 0.02318406742965053$	4.5	$0.423 \pm 0.03063595461733346$
5	$0.19 \pm 0.024327228655439855$	5	$0.426 \pm 0.030664371868702973$

D.4 Modell 4 - Modell för concertillväxt

I Figur D.1 visas hur fixeringssannolikhet beror på bärformåga respektive reproduktionstakt för antal gitterpunkter $L = 10$, samt skillnaden för extremvärdena på specificiteten, γ . Bärformågan varierar enligt $k_A = \{1, 2, 3, 4, 5\}$ och reproduktionstakten varierar enligt $r_A = \{1, 1.2, 1.4, 1.6, 1.8, 2\}$.



Figur D.1: Illustration av skattat väntevärde samt 95 % konfidensintervall av fixeringssannolikhet för cancertillväxt när $L = 10$ plottat för specificiteterna $\gamma = 0$, som är de röda punkterna, och $\gamma = 1$, som är de gröna punkterna. Parametervärdena för de olika figurerna var **(a)** $r_A = 1$ och varierande k_A , och **(b)** $k_A = 1$ och varierande r_A .

Följande tabeller visar skattat väntevärde samt tvåsidiga 95 % konfidensintervall för fixeringssannolikhet för cancertillväxt. Varje fall testade extremvärdena för specificiteten γ , det vill säga $\gamma = 0$ och $\gamma = 1$. I Tabellerna D.5 och D.6 beskrivs intervallen för varierande reproduktionstakt hos genotyp A, $r_A = \{1.0, 1.2, 1.4, 1.6, 1.8, 2.0\}$ för antalet platser $L = 10$ respektive $L = 100$. I Tabell D.7 och D.8 beskrivs intervallen för varierande bärförmåga hos genotyp A, $k_A = \{1, 2, 3, 4, 5\}$ för antalet platser $L = 10$ respektive $L = 100$.

Tabell D.5: Skattat väntevärde samt tvåsidigt 95 % konfidensintervall för fixeringssannolikhet P_f för Modell 4 med varierande reproduktionstakt r_A och konstant antal gitterpunkter $L = 10$ för specificitetskonstanterna **(a)** $\gamma = 0$ och **(b)** $\gamma = 1$.

(a)		(b)	
r_A	I_{P_f}	r_A	I_{P_f}
1	$0.089 \pm 0.017657431049596974$	1	$0.088 \pm 0.017567585766680337$
1.2	$0.205 \pm 0.025034206028381877$	1.2	$0.193 \pm 0.024473086719365284$
1.4	$0.272 \pm 0.027594559705226774$	1.4	$0.264 \pm 0.027334692824903514$
1.6	$0.323 \pm 0.028998058572963142$	1.6	$0.306 \pm 0.028576811868656314$
1.8	$0.352 \pm 0.02961639283107118$	1.8	$0.34 \pm 0.029375465238153382$
2	$0.39 \pm 0.030246181105578793$	2	$0.41 \pm 0.03049939758836645$

Tabell D.6: Skattat väntevärde samt tvåsidigt 95 % konfidensintervall för fixeringssannolikhet P_f för Modell 4 med varierande reproduktionstakt r_A och konstant antal gitterpunkter $L = 100$ för specificitetskonstanterna **(a)** $\gamma = 0$ och **(b)** $\gamma = 1$.

(a)		(b)	
r_A	I_{P_f}	r_A	I_{P_f}
1	$0.007 \pm 0.005170076526637811$	1	$0.009 \pm 0.005856409132559638$
1.2	$0.162 \pm 0.02284824483175659$	1.2	$0.162 \pm 0.02284824483175659$
1.4	$0.223 \pm 0.02581286845311419$	1.4	$0.231 \pm 0.026136202545195497$
1.6	$0.32 \pm 0.02892695851500688$	1.6	$0.31 \pm 0.028679971770920222$
1.8	$0.334 \pm 0.029247158266516537$	1.8	$0.335 \pm 0.02926891028222795$
2	$0.392 \pm 0.030273884548991818$	2	$0.377 \pm 0.030053013446528408$

Tabell D.7: Skattat väntevärde samt tvåsidigt 95 % konfidensintervall för fixeringssannolikhet P_f för Modell 4 med varierande bärförmåga k_A och konstant antal gitterpunkter $L = 10$ för specificitetskonstanterna (a) $\gamma = 0$ och (b) $\gamma = 1$.

(a)		(b)	
k_A	I_{P_f}	k_A	I_{P_f}
1	$0.112 \pm 0.019556395941538466$	1	$0.086 \pm 0.017385838886685732$
2	$0.264 \pm 0.027334692824903514$	2	$0.896 \pm 0.01892971178830751$
3	$0.308 \pm 0.028628706929591848$	3	$0.92 \pm 0.01682334047639721$
4	$0.328 \pm 0.029113531275049807$	4	$0.897 \pm 0.018848993345803648$
5	$0.366 \pm 0.02987160094375497$	5	$0.916 \pm 0.017201277695402577$

Tabell D.8: Skattat väntevärde samt tvåsidigt 95 % konfidensintervall för fixeringssannolikhet P_f för Modell 4 med varierande bärförmåga k_A och konstant antal gitterpunkter $L = 100$ för specificitetskonstanterna (a) $\gamma = 0$ och (b) $\gamma = 1$.

(a)		(b)	
k_A	I_{P_f}	k_A	I_{P_f}
1	$0.008 \pm 0.005524260588988816$	1	$0.016 \pm 0.0077809185250260204$
2	$0.253 \pm 0.026958386082630764$	2	$0.88 \pm 0.02015140290498503$
3	$0.276 \pm 0.02772025055245155$	3	$0.895 \pm 0.019009885211008923$
4	$0.361 \pm 0.029783611431623096$	4	$0.898 \pm 0.01876772285124173$
5	$0.351 \pm 0.02959710504470035$	5	$0.895 \pm 0.019009885211008923$

I Tabell D.9 presenteras värdena för fixeringssannolikhet för varierande bärförmåga och specificitet enligt $k_A = \{1, 2, 3, 4, 5\}$ respektive $\gamma = \{0, 0.25, 0.5, 0.75, 1\}$. Dessa värden är rådata som användes för illustrationen i Figur 4.7 i resultatavsnittet.

Tabell D.9: Värden för fixeringssannolikhet för Modell 4 med varierande specificitetskonstant γ och bärförmåga k_A för $r_A = 1$.

		Bärförmåga k_A				
		1	2	3	4	5
Specificitet γ	0	0.098	0.275	0.329	0.363	0.359
	0.25	0.095	0.786	0.861	0.87	0.895
	0.5	0.114	0.848	0.904	0.89	0.898
	0.75	0.116	0.872	0.892	0.896	0.917
	1	0.089	0.886	0.881	0.907	0.908

I Tabell D.10 och D.11 finns värden för fixeringssannolikhet för varierande bärförmåga och reproduktionstakt enligt $k_A = \{1, 2, 3, 4, 5\}$ respektive $r_A = \{1, 1.2, 1.4, 1.6, 1.8, 2\}$. Dessa värden är rådata som användes för illustrationerna i Figur 4.6 i resultatavsnittet.

Tabell D.10: Värderna för fixeringssannolikhet för Modell 4 med varierande reproduktionstakt r_A och bärförmåga k_A för $\gamma = 0$.

		Bärförmåga k_A				
		1	2	3	4	5
Reproduktionstakt r_A	1.0	0.095	0.248	0.302	0.342	0.362
	1.2	0.187	0.375	0.429	0.448	0.518
	1.4	0.27	0.45	0.508	0.582	0.588
	1.6	0.295	0.504	0.571	0.597	0.627
	1.8	0.344	0.551	0.607	0.659	0.708
	2.0	0.393	0.603	0.606	0.69	0.741

Tabell D.11: Värderna för fixeringssannolikhet för Modell 4 med varierande reproduktionstakt r_A och bärförmåga k_A för $\gamma = 1$.

		Bärförmåga k_A				
		1	2	3	4	5
Reproduktionstakt r_A	1.0	0.109	0.891	0.888	0.916	0.892
	1.2	0.154	0.901	0.909	0.922	0.919
	1.4	0.257	0.914	0.931	0.919	0.913
	1.6	0.281	0.935	0.936	0.929	0.919
	1.8	0.36	0.935	0.933	0.943	0.939
	2.0	0.41	0.941	0.953	0.961	0.939