



# Galton-Watson-processen och åldrande celler

Matematisk modell av åldrande och förnyring

The Galton-Watson-process and aging cells

*Kandidatarbete inom civilingenjörsutbildningen vid Chalmers*

Lotta Eriksson

Joel Hildingsson

Ibrahim Taha

Laleh Varghaei



# Galton-Watson-processen och åldrande celler

Matematisk modell av åldrande och förnygring

*Kandidatarbete i matematik inom civilingenjörsprogrammet Bioteknik vid Chalmers*  
Joel Hildingsson   Laleh Varghaei

*Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik vid Chalmers*  
Lotta Eriksson   Ibrahim Taha

Handledare:   Serik Sagitov

Institutionen för Matematiska vetenskaper  
CHALMERS TEKNISKA HÖGSKOLA  
GÖTEBORGS UNIVERSITET  
Göteborg, Sverige 2022



## Förord

Gruppen vill rikta ett stort tack till vår handledare Serik Sagitov för handledning och givande diskussioner under kursens gång. Gruppen vill även tacka ämnesexpert Marija Cvijovic för input kring den biologiska tillämpningen av projektet och kring forskningen som bedrivs i hennes forskargrupp.

Vi vill även tacka alla som har läst och gett respons på arbetet under kursens gång: medlemmarna i MVEX01-22-03 under det första handledningstillfället, Hans Malmström under andra handledningstillfället och Victor Salomonsson.

## Bidragsrapport

I tabellen nedan redovisas vilka delar av rapporten som skall tillskrivas vilken gruppmedlem. Kolumnen *Författare* anger vem som i huvudsak ansvarat för texten. Om någon annan medlem bidragit med text till ett avsnitt tillskriven en annan medlem noteras detta längst ned på sidan.

§	Rubrik	Författare
	Populärvetenskaplig rapport	Joel
	Sammanfattning	Lotta/Joel
	Abstract	Lotta/Joel
	Notationslista	Alla
1.	<b>Inledning*</b>	
1.1	Syfte	Ibrahim
1.2	Problemformulering	Ibrahim
1.3	Avgränsningar	Lotta
1.4	Samhälleliga och etiska aspekter	Ibrahim
2.	<b>Biologisk bakgrund</b>	
2.1	Olika åldrande i celler	Laleh
2.2	Bromsande av biologisk ålder genom föryngring	Laleh
2.3	Egenskaper hos cellpopulationer	Lotta
2.4	Extrema åldrar hos människor	Laleh
3.	<b>Matematisk modell av åldrande celler**</b>	Lotta
4.	<b>Metod</b>	Ibrahim
5.	<b>Resultat</b>	
5.1	Kritisk gräns för utrotning	Lotta
5.2	Fördelning av biologiska åldrar och populationsstorlek	Lotta
5.3	Cellers levnadslängdsfördelning***	Ibrahim
5.4	Levnadslängdsfördelningens svans följer en geometrisk fördelning	Laleh
5.5	Föryngring inom populationer	Lotta
6.	<b>Diskussion</b>	
6.1	Tolkning av resultat****	Alla
6.2	Modellens begränsningar och framtida forskning	Joel
6.3	Avslutning	Lotta
	Programkod*****	Lotta

\*Ibrahim har haft huvudansvar för inledningen. Laleh skrev första stycket i inledningen. Lotta skrev ursprungligen ett avsnitt om den enkla Galton-Watson-processen som sedan togs bort som skrevs om och lades istället in i bakgrunden. Ibrahim har sedan skrivit ihop de olika delarna.

\*\*Matematisk modell och de teoretiska resultaten specifika för modellen är skrivna av Lotta. Samtliga medlemmar har bidragit avsnittet i form av teori kring flertyp Galton-Watson-processen. Detta avsnitt togs emellertid bort i slutrapporten och skrevs ihop med matematisk modell av Lotta.

\*\*\* Ibrahim har haft huvudansvar för cellers levnadslängdsfördelning. Laleh skrev om påverkan av cellens initial biologisk ålder.

\*\*\*\*Tolkning av resultat är skriven av Ibrahim och Laleh. Samtliga medlemmar har bidragit i form av analyser och slutsatser för tolkning av resultatet. Avsnittet slogs ihop i slutrapporten.

\*\*\*\*\*Joel gjorde en prototyp till simulering av åldersfördelningen i `Python` som sedan översattes till `C++` av Lotta för minska beräkningstid och tolka datan i `R`. Joel och Lotta gjorde koden för föryngringsindex i `C++`.

### **Gruppens arbete**

Under arbetets gång har en loggbok förts över medlemmarnas individuella prestationer, dels i en veckovis dagbok men även i individuella tidsloggar. Arbetet att skriva dagbok har fördelats lika mellan samtliga medlemmar genom att ansvaret roterats varje vecka. Under arbetets gång har gruppen haft veckovisa möten med handledaren samt i genomsnitt två möten med endast gruppen. Samtliga medlemmar har deltagit vid dessa moment. Alla medlemmar i gruppen har bidragit till texten genom att kontinuerligt jobbat med att läsa, kommentera och bearbeta varandras delar i texten.

## Populärvetenskaplig presentation

I naturen dör organismer ofta från skador, rovdjur, sjukdom eller brist på näring. De flesta får inte leva ut sin livslängds fulla potential. Hos människor är detta möjligtvis inte fallet längre, allt fler dör istället från sjukdomar som orsakas av skador som sakta ackumulerats under deras livstid. Dessa skador ökar en individs biologiska ålder vilket kan ses som ett mått på hur frisk den är. En typ av skador som ackumuleras är felveckade eller oxidativt skadade proteiner, dessa kan leda till sjukdomar som till exempel Alzheimers [1, s. 71]. Att modellera åldrande är svårt då det är ett så komplicerat fenomen och det är därför av stort intresse att hitta enklare modeller.

Mycket av arbetet jämförs med beteenden hos *Saccharomyces cerevisiae*, bagerijäst. Man ser vissa likheter i resultaten med deras beteende men det skiljer sig även på vissa håll. Jäst är en lämpad organism att jämföra med eftersom det är en välstuderad modellorganism och vi har även tagit inspiration av andra arbeten som studerat åldrandeprocessen hos jäst. Då jästceller är eukaryoter kan ny kunskap om deras beteenden leda till slutsatser om andra eukaryoter som människor eller andra djur.

Galton-Watson-processen utvecklades av Francis Galton tillsammans med H. W. Watson som tillsammans undersökte varför adliga efternamn dog ut i England. Deras resultat publicerades som 'On the Probability of the Extinction of Families' året 1875 [2, s. 50]. Sedan dess har Galton-Watson-processen tillämpats bland annat inom genetik och ekologi [3, preface]. Här kommer Galton-Watson-modellen istället att användas i ett försök att modellera olika typer av åldrande. Vi kommer att använda sig av biologiskt åldrande, som innebär hur gammal man är i praktiken och man tittar på kronologisk ålder som är tiden man har levt. Två individer med samma kronologiska ålder kan ha olika biologiska ålder beroende på faktorer så som genetik och livsstil.

Modellen går ut på att simulerade celler har en gräns med maximalt antal defekta proteiner som de tolererar och de ackumulerar ett fast antal defekta protein varje tidsenhet. Cellerna har en sannolikhet att dela sig i varje tidssteg och när de delar sig ärver dottercellen ett slumpmässigt antal defekta proteiner. Beroende på sannolikheten för fördelningen av proteiner kan celledelningen klassificeras som asymmetriskt eller symmetrisk. Vid asymmetrisk fördelning kommer modern behålla de flesta protein i sig själv och därmed skydda dottern från att födas biologiskt gammal. Vid symmetrisk fördelning delas de defekta proteinerna upp jämnt mellan moder och dotter.

Beroende på hur dessa egenskaper förhåller sig till varandra kan man se om en population garanterat kommer att bli utrotad. Det går även att hitta fördelningen av biologiska åldrar inom en population och fördelningen av livslängder för olika typer av celler. Inspirerat av andra arbeten undersöker man även svansen av livslängdsfördelningen. Vi studerar också fenomenet föryngring, i detta fall är föryngring ett mått på hur livslängd ökar eller minskar mellan generationer.

Även om modellen har sina brister har den lyckats fånga in vissa aspekter av jästcellers beteende. Detta visar på att även enkla modeller med få parametrar kan ha möjligheten att härma beteenden i större och mer komplicerade system.

## Sammanfattning

Teorin kring den flertypa Galton-Watson-processen har tillämpats för att skapa en matematisk modell som beskriver biologiskt åldrande hos celler. Vi studerar även fenomenet föryngring. Utifrån den matematiska modellen tillämpas teoretiska resultat och populationer simuleras för att studera hur biologiskt åldrande påverkar cellpopulationer och de individuella cellerna.

I arbetet undersöks hur modellens olika parametrar påverkar egenskaperna hos cellpopulationerna. Vi studerar fördelningen av olika biologiska åldrar i populationen, garanterad utrotning av populationer och levnadslängder hos celler. Vi introducerar även ett föryngringsindex för att ge ett formellt mått på föryngring i populationer.

Det är i huvudsak ackumulationshastigheten av skador och hur frekvent cellerna delar sig som påverkar huruvida populationen utrotas. I de överlevande populationerna resulterar asymmetrisk fördelning av defekta proteiner mellan moder- och dottercell i mindre populationer men en högre andel biologisk unga celler. De individuella cellerna lever under en kortare tid men föryngring sker i högre grad. I populationer där skadan fördelas symmetriskt blir populationerna större och de individuella cellerna lever längre men föryngring sker inte.

Den slutgiltiga modellen har på ett enkelt men insiktsfullt sätt modellerat åldrandet hos celler men det finns en viss känslighet i valet av parametrar. Den matematiska modell som utvecklats har även jämfört åldrandet och föryngring med jästcellen. Framtida forskning syftar till justering av modellens parametrar för att kunna beskriva mekanismerna bakom åldrandet hos jästcellen och få en insikt i det mänskliga åldrandet.

## Abstract

The theory of the multitype Galton-Watson-process has been applied to build a mathematical model for describing biological aging in cells. We also study the phenomenon of rejuvenation. The theoretical results of the model have been applied and populations have been simulated in order to study how biological aging affects cell populations and individual cells.

This project investigates how different parameters of the model affect the properties of cell populations. We study what the distribution of different biological ages looks like in the population, guaranteed extinction of populations and the distribution of life lengths in cells. We also introduce a rejuvenation index to provide a formal measure of rejuvenation in the populations.

The rate of accumulation of damage and how frequently the cells divide is essential in whether or not the population becomes extinct. In the surviving populations, asymmetric distribution of damage between the mother and daughter cells results in smaller populations but with a higher proportion of biologically young cells. The individual cells live for a shorter period of time but rejuvenation occurs to a higher extent. In populations where the damage is distributed symmetrically, populations become larger and the individual cells live longer but rejuvenation does not occur.

Aging in cells has been modeled in a simple but insightful way but there is still some sensitivity in the choice of parameters. The mathematical model that has been developed has also compared the aging and rejuvenation with the yeast cell. We suggest that future research aims at adjusting the parameters of the model in order to be able to describe the mechanisms behind the aging of the yeast and gain an insight into human aging.



## Notation

$e_i$	Enhetsvektorn $(0, \dots, 1, \dots, 0)$ med 1 på position $i$
$\mathbb{N}_0$	De naturliga talen med 0 inkluderat $\{0, 1, \dots\}$
$\mathbf{1}\{P\}$	Indikatorfunktionen med värde 1 om påståendet $P$ är sant och 0 annars
$\Pr(A)$	Sannolikheten för en händelse $A$ .
$\mathbb{E}[X]$	Väntevärdet av slumpvariabeln $X$
$Z_t^{(i)}$	Antalet individer av typ $i$ vid tidpunkt $t$
$\mathbf{Z}_t$	Vektor med antalet individer för flertyp Galton-Watson-process vid tidpunkt $t$
$\mathbf{M}$	Väntevärdesmatrisen i flertyp Galton-Watson-process
$\rho$	Perron-Frobenius egenvärdet till $\mathbf{M}$
$\mathbf{u}$	Höger egenvektor till $\mathbf{M}$ tillhörande egenvärdet $\rho$
$\mathbf{v}$	Vänster egenvektor till $\mathbf{M}$ tillhörande egenvärdet $\rho$
$g$	Sannolikheten för ett lyckat utfall i den geometriska fördelningen

# Innehåll

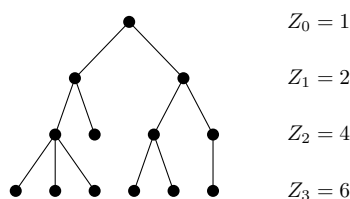
<b>1 Inledning</b>	<b>1</b>
1.1 Syfte . . . . .	2
1.2 Problemformulering . . . . .	2
1.3 Avgränsningar . . . . .	2
1.4 Samhälleliga och etiska aspekter . . . . .	2
<b>2 Biologisk bakgrund</b>	<b>3</b>
2.1 Olika åldrande i celler . . . . .	3
2.2 Bromsande av biologisk ålder genom föryngring . . . . .	3
2.3 Egenskaper hos cellpopulationer . . . . .	4
2.4 Extrema åldrar hos människor . . . . .	4
<b>3 Matematisk modell av åldrande celler</b>	<b>5</b>
3.1 Väntevärdesmatrisen . . . . .	5
3.2 Garanterad utrotning av populationen . . . . .	6
3.3 Stabila typfördelningen och en populations hälsa . . . . .	7
3.4 Egenskaper på populations- och individnivå . . . . .	7
3.5 Specialfall av modellen . . . . .	7
<b>4 Metod</b>	<b>8</b>
<b>5 Resultat</b>	<b>8</b>
5.1 Kritisk gräns för utrotning . . . . .	8
5.2 Fördelning av biologiska åldrar och populationsstorlek . . . . .	9
5.3 Cellers levnadslängdsfördelning . . . . .	11
5.4 Levnadslängdfördelningens svans följer en geometrisk fördelning . . . . .	13
5.5 Föryngring inom populationer . . . . .	13
<b>6 Diskussion</b>	<b>15</b>
6.1 Tolkning av resultat . . . . .	15
6.2 Modellens begränsningar och framtida forskning . . . . .	16
6.3 Avslutning . . . . .	16
<b>Referenser</b>	<b>18</b>
<b>Appendix</b>	<b>19</b>
<b>A Övriga resultat</b>	<b>19</b>
A.1 Geometrisk fördelning och initial biologisk ålder . . . . .	19
<b>B Simulering och programkod</b>	<b>19</b>
B.1 Programkod . . . . .	19
B.2 Simulering . . . . .	19
<b>C Grundläggande sannolikhetsteori</b>	<b>19</b>
C.1 Stokastiska processer och Markovkedjor . . . . .	19
C.2 Sannolikhetsfördelningar . . . . .	20
C.3 Maximum likelihood för geometrisk fördelning . . . . .	20

# 1 Inledning

Under de senaste decennierna har intresset för att förstå åldrande vuxit snabbt inom forskning. Idag har människor längre medellivslängd, däremot ökar åldersrelaterade sjukdomar i takt med att medellivslängden ökar. Detta ger en anledning till att vilja förstå mekanismen bakom processen för åldrande. Denna förståelse kan bidra till att designa experiment för att förstå vilka sjukdomsalstrande förändringar som sker i kroppen i samband med åldrandet och att analysera dessa förändringar på molekylär och cellulär nivå. Informationen som fås från experiment och forskning på cellulär- och molekylär nivå kan hjälpa forskare att förstå bland annat hur de kan förebygga eller bota åldersrelaterade sjukdomar.

Alzheimers är ett exempel på en åldersrelaterad sjukdom som uppstår på grund av ackumulation av defekta proteiner. Sjukdomen kan inom biologi beskrivas med hjälp av slumpprocesser [1]. Ideén är att defekta protein inom celler fördelas enligt en slumpprocess och därmed styr cellpopulationer. Dessa protein kan därmed orsaka utrotning för hela populationen. Ackumulation av defekta protein är en slumpprocess som har blivit allt mer populär. För att modellera sådana sorters processer och förstå dem bättre används Galton-Watson-processer.

Galton-Watson-processen är en stokastisk förgreningsprocess som har flera användningsområden inom många olika fält. Modellen är uppdelad i två sorter, enkel och flertyp. Processen uppkom ursprungligen för att studera huruvida efternamn dör ut över tid, vilket var viktigt för aristokratin i Storbritannien under 1800-talet. Galton-Watson-processen kan även tillämpas på mer allmänna problem där populationer av icke urskiljbara partiklar studeras över tid. Modellen har fått särskilt genomslag i biologisk forskning och används bland annat för att studera spridning av smittsamma sjukdomar i samhället över tid [4]. Idag är modellen mer avancerad och används därmed inom mer komplexa områden, bland annat populationers beteende utifrån olika faktorer, så som defekta protein.



Figur 1: Visualisering av tre generationer av Galton-Watson-processen där  $Z_t$  anger antalet individer i generation  $t$ .

Processen bygger på att individer oberoende av varandra föder ett slumpmässigt antal individer enligt en sannolikhetsfördelning som kallas för *reproduktionsfördelningen*. Det förväntade antalet individer från en individ kallas för *reproduktionsvärdet* och styr huruvida populationen garanterat kommer att dö ut eller ha en positiv sannolikhet att överleva [4, s. 160-161]. Utifrån storleken på reproduktionsvärdet klassificeras processen i olika kategorier med olika långsiktigt beteende. Om reproduktionsvärdet är mindre än eller lika med 1 garanteras utrotning av populationen och kallas subkritisk respektive kritisk. Om väntevärdet är större än 1 finns en positiv sannolikhet att populationen aldrig utrotas med förväntad exponentiell tillväxt [4]. Tre generationer av Galton-Watson-processen visualiseras i Figur 1.

För att modellera mer komplexa biologiska system där individerna även kan vara av olika typer används istället en flertyp Galton-Watson-process [5]. Ett exempel på tidigare forskning som använder flertyp Galton-Watson-processer är fortplantningen av parasiter bland cellpopulationer, där mängden parasiter bland celler följer en slumpmässig förgreningsprocess. Dessa parasiter har förmågan att från modercellen slumpmässigt överförs till dottercellerna. En konsekvens av detta är att om antalet parasiter i en cell överstiger en bestämd gräns, leder detta till celledöd.

Ett annat sätt att modellera åldrande celler är genom att använda ODE:er [3, s.65-66], vilket detta arbete inte tar del av. Istället används en flertyps Galton-Watson-process för att modellera

de åldrande cellerna och studera populationen. De biologiska processerna för åldrande celler som ackumulerar defekta protein är för komplexa för att fullständigt täckas ur ett rent biologiskt perspektiv. De matematiska modellerna, i form av stokastiska processer, klargöra detta då de klarar av att hantera sådana komplexa moment. Därför är Galton-Watson-processen en lämplig modell för sådana problem. Problemet i detta arbete som undersöks är åldrande cellers population samt defekta proteins påverkan på populationen. Dessa defekta protein har förmågan att ackumuleras och orsaka celldöd.

## 1.1 Syfte

Projektets syfte är att bygga upp en modell, beskriva samt simulera en population av åldrande celler under diskret tid med hjälp av Galton-Watson-processen. Cellerna åldras under förutsättningen att defekta protein ackumuleras. Detta genomförs genom att tillämpa analytiska metoder samt datorsimuleringar.

Arbetet genomförs även för att undersöka vilken roll defekta protein har för de åldrande cellerna, medellivslängden för celler i varje population, samt hur föryngring ser ut för cellers ålder. Utifrån detta vill vi undersöka fördelningen av celler med låga antal defekta proteiner i levande populationerna jämfört med celler som har defekta protein.

## 1.2 Problemformulering

Målet med arbetet är att besvara följande frågeställningar:

- Vilka parametervärden säkerställer att populationen dör ut?
- Hur ser en stabil typ-fördelning ut i stora populationer?
- Hur ser fördelningen av möjliga medellivslängder för moderceller ut?
- Påverkar föryngring populationstillväxten?

## 1.3 Avgränsningar

Projektets syfte är studera en matematisk modell av biologiskt åldrande celler, inte att studera biologiskt åldrande celler i en verklig miljö. Vi kommer således inte att studera och samla in data från celler i verkliga system utan undersökningen görs analytiskt och med simuleringar. Det finns inte heller någon given datamängd som kommer studeras under arbetet. Vidare begränsar vi den teoretiska modellen genom att endast studera diskret tid, inte kontinuerlig.

## 1.4 Samhälleliga och etiska aspekter

Ur ett etiskt perspektiv, innehåller arbetet inga etiska dilemman eller synpunkter som tas hänsyn till då det inte hanterar verkliga biologiska celler, utan simulerar rent matematiska celler. Dessutom samlas inte data in från andra källor, utan dessa skapas genom simuleringarna. Från en samhälls synpunkt bidrar arbetet med nytt sätt att modellera åldrande celler på samt resultat för forskare som arbetar inom matematisk biologi. Detta ger en mer matematisk inblick över ämnen som celldöd, ackumulation av protein och populationsbildning, vilket kan bidra med ett nytt sätt att modellera biologiskt åldrande och föryngring.

## 2 Biologisk bakgrund

Åldrande är en komplex fysiologisk process. Det kan antas vara en multifaktoriell process som påverkas av miljön som organismen eller arten lever i, men i allmänhet definieras åldrande som vad som händer i en organism över tiden [1, s. 3] [6, s. 44]. Åldrande klassificeras i olika kategorier så som biologiskt, kronologiskt eller replikativt åldrande [6]. I det här kapitlet definieras olika typer av åldrande, föryngring i celler och föryngringsindex. Föryngring är en mekanism som bromsar det biologiska åldrandet [7]. Kapitlet avslutas med en diskussion om extrema åldrar hos människor.

### 2.1 Olika åldrande i celler

Man kan tala om olika sorters av åldrande i celler och en av dem är biologiskt åldrande. De biologiska processerna i kroppen är resultatet av olika aktiviteter på den molekylära nivån. Biologiskt åldrande kan därmed definieras som ackumulering av skador på cellulär och molekylär nivå. Dessa skador påverkar funktionen hos celler och organismer och i slutändan hela kroppen. En av de skador som leder till ökat biologiskt åldrande är ansamling av felveckade proteiner i celler [8] [1, s. 3].

Proteiner produceras i celler genom translationsprocessen. Kodon i messenger RNA läser av ribosomer i translationsprocessen och det leder till produktion av aminosyrasekvenser. Protein är långa aminosyrasekvenser. Dessa proteiner ska vikas korrekt för att få en specifik 3D-struktur. 3D-strukturer gör det möjligt för proteiner att vara biologiskt aktiva. Ibland kan dessa proteiner vikas inkorrekt och få en felaktig 3D-struktur vilket kallas felveckade proteiner. Det finns många faktorer som orsakar felveckning så som mutationer eller oxidativ stress. Celler har förmågan för att bryta ned de felveckade proteinerna men ibland misslyckas nedbrytningen vilket leder till ackumulering av felveckade proteiner i cellen. Denna ansamling utgör en allvarlig skada för cellen och ökar cellens biologiska ålder eller kan till och med leda till celledöd [9, s. 340-361]. Ackumulerade felveckade proteiner i celler kan orsaka neurodegenerativa sjukdomar som Alzheimers [1, s. 71].

För att definiera det kronologiska åldrandet tar man bara hänsyn till den tid som organismen lever från födelsen. Bland olika arter och celler finns stor variation inom livslängd. Till exempel lever röda blodkroppar, som bär syre i våra kärl, i cirka fyra månader medan slemceller har en livslängd på bara några dagar. De celler som lever längst är neuronerna som omvandlar signaler i nervsystemet och har en mycket låg regenereringsförmåga [1, s. 7]. Celler som lever längre, har även längre tid på sig att ackumulera skador vilket gör dem särskilt utsatta för skaderisk över tid. Ackumulerade skador i neuroner kan orsaka oåterkalleliga tillstånd i hjärnan och ökad kronologisk ålder ökar risken för skador [1].

Det biologiska och kronologiska åldrandet är i själva verket inte samma sak och en cell eller en organism kan ha olika biologisk respektive kronologisk ålder samtidigt [8]. Förutom biologiskt och kronologiskt åldrande finns det en annan typ av åldrande som kallas replikativt åldrande. Replikativt åldrande i celler definieras som antalet celledelningar av en modercell under modercellens livslängd. Reproduktion är ett viktigt steg i varje cells livslängd. Med hjälp av replikativt åldrande kan man observera hur ackumulerade skador, mutationer och även tillämpning av mediciner påverkar en cells reproduktionsförmåga över tiden [10] [6, s. 80].

### 2.2 Bromsande av biologisk ålder genom föryngring

Föryngring är en process i en organism som vänder eller bromsar det biologiska åldrandet. Föryngringsprocessen betraktas som motsatsen till åldrande. I denna process försöker organismen reparera eller ta bort den ackumulerade skadan över tid och det leder till en förbättring av den funktionella nivån hos en organism som skadats över tid på grund av åldrande. På det här sättet ökar organismens kronologiska och replikativa ålder, samtidigt som den biologiska åldern minskar [10] [7].

Föryngring kan ske vid celledelning. När en modercell ackumulerar skador, till exempel ackumulering av felveckade proteiner, kan dessa skador överföras till dotterceller asymmetriskt. Dottercellen kan ta över skadorna och biologiskt sett födas gamla på grund av de ackumulerade skadorna trots

att dottercellens kronologiska ålder är låg. I detta fall kan föryngring definieras som att behålla skadorna i modercellen och föda friska barn. Man kan säga att dottercellen har föryngrats och dess biologiska ålder minskat trots att den kronologiska åldern ökar med tiden [10] [6, s. 80].

Föryngring och åldrande kan studeras i jästen *S. cerevisiae*. Resultat från forskning på den här organismen som är en encellig eukaryot kan appliceras på flercelliga eukaryoter såsom människor [6, s. 80] [9, s. 31]. Celldelningen i denna jäst är asymmetrisk och kallas knoppning. Vid celldelningen dyker en liten knopp upp på modercellens kropp och utvecklas efter ett tag som en ny oberoende cell. Dottercellen är mindre än modercellen i storlek och överföringen av skadade proteiner sker asymmetrisk mellan modern och dottern. Det finns såväl passiva som aktiva mekanismer som försöker behålla de skadade proteinerna i modercellen och förhindra att dessa skador överförs till nästa generation. En av de passiva mekanismerna är den smala knoppalsen som gör det svårt för skadade proteiner att överföras till den nya dottercellen. Modercellen offerar sig själv genom att behålla skador och föder därmed biologiskt unga dotterceller. Denna föryngring ger upphov till en friskare population, den möjliggör högre replikativ ålder hos dotterceller och hela populationens livslängd blir längre [10] [9, s. 925].

### 2.3 Egenskaper hos cellpopulationer

I tidigare studier har åldrandeprocessen och föryngring i jästcellen *S. cerevisiae* modellerats [10]. På populationsnivå har egenskaper såsom populationsstorlek, andel hälsosamma celler och andelen föryngrade celler studerats. För att studera ett formellt mått på graden av föryngring har ett så kallat *föryngringsindex re* introducerats som anger skillnaden i dotter- och modercellen replikativa ålder skalat med medellivslängden i populationen. Arbetet definierar föryngring som att dotterceller får högre replikativ ålder än modercellen och sker då  $re > 0$ . Med inspiration från tidigare studier studerar vi i detta arbete liknande egenskaper hos cellpopulationer men med en flertyp Galton-Watson-process som modell.

### 2.4 Extrema åldrar hos människor

Syftet med utveckling av läkemedel och all medicinsk forskning är att förbättra människors livskvalitet och att förlänga livslängden. Det finns inget självklart svar ännu om det finns en gräns för människans livslängd och det är en kontroversiell diskussion bland forskarna [11].

Även om den förväntade livslängden för människor har ökat mycket under de senaste decennierna, har ökningen i livslängd för de som blir över 110 år varit relativt långsam. Ibland kallas personer som uppnår denna ålder för supercentenärer. Att studera livslängden hos supercentenärer kan hjälpa oss att förstå mekanismen bakom åldrandet ännu bättre [11].

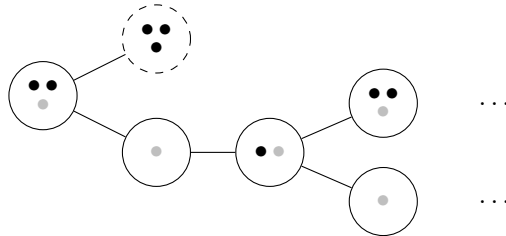
Holger Rootzén och Dmitrii Zholud hävdar, i artikeln 'Human life is unlimited – but short', att sannolikheten att en person som har fyllt 110 år ska leva ytterligare ett år är 50 procent. Denna sannolikhetsbeskrivning är som att singla slant och om till exempel krona kommer upp kan personen leva ett år till [11]. Att singla slant anses vara geometriskt fördelat. Med hjälp av geometrisk fördelning, som är en diskret fördelning, kan antalet försök beräknas innan det första lyckade utfallet inträffar [12, s. 114]. Sannolikhet för ett lyckat utfall i en geometrisk fördelning betecknas med  $g$  i vårt projekt.

Enligt artikeln ovan är sannolikheten att en kvinna lever i mer än 110 år tio gånger mer än en man. Även om den förväntade livslängden och överlevnaden hos människor i allmänhet beror på faktorer som land, kön, livsstil osv, spelar dessa faktorer ingen roll i dödligheten för supercentenärerna. Det finns inte något självklart svar ännu eller en fullständig konsensus i forskarvärlden, på om det finns en gräns för människans livslängd. Däremot finns det resultat som pekar på att det kanske faktiskt inte finns någon absolut gräns. Även om människans liv kan uppfattas kort [11].

### 3 Matematisk modell av åldrande celler

För att modellera och studera populationer av biologiskt åldrande celler använder vi oss av en tidsdiskret stokastisk förgreningsprocess. Modellen som används är en tillämpning av en flertyp-Galton-Watson-process. Cellerna i populationen åldras biologiskt genom att defekta proteiner ackumuleras över tid. Antalet defekta proteiner gör cellen till en specifik typ  $\{0, 1, \dots, k\}$  och motsvarar dess biologiska ålder. Oberoende av varandra föder modercellen med viss sannolikhet en ny dottercell i varje tidssteg och fördelar proteinerna mellan moder- och dottercell enligt en sannolikhetsfördelning. Förgreningsprocessen påbörjas alltid i en cell av en specifik typ. Modellen har parameteruppsättningen  $(p, n, k, q)$  där:

- $p$ : sannolikheten att ett defekt protein ärvs av dottercellen vid celledelning
- $n$ : antalet nya defekta protein som ackumuleras varje tidsenhet
- $k$ : det maximala antalet defekta protein cellen kan ha utan att dö
- $q$ : sannolikheten för att en modercell producerar en dottercell varje tidssteg



Figur 2: Visualisering av modellen med  $(n, k) = (1, 2)$ . De svarta cirklarna ( $\bullet$ ) motsvarar antalet defekta proteiner och de ljusgråa cirklarna ( $\circ$ ) representerar de nya defekta proteiner som ackumuleras av cellen varje tidsenhet. Förgreningsprocessen påbörjas i en cell med två defekta proteiner. I första tidssteget delar sig cellen och dottercellen ackumulerar inga defekta proteiner. Modercellen har nu  $3 > k$  defekta proteiner och dör. Dottercellen överlever och delar sig inte i nästa tidssteg och ackumulerar alla defekta proteiner. I det tredje tidssteget delar sig cellen och både moder- och dottercellen överlever. Förgreningsprocessen fortskrider sedan på samma sätt.

Vi låter  $\mathbf{Z}_t = (Z_t^{(0)}, \dots, Z_t^{(k)})$  vara en  $k + 1$ -dimensionell vektor där  $Z_t^{(i)}$  anger antalet celler med  $i$  defekta protein vid tidpunkten  $t$ . Låt vidare  $\{\mathbf{Z}_t\} = (\mathbf{Z}_0, \mathbf{Z}_1, \dots)$  vara sekvensen av tillstånd. Vi antar genom arbetet att  $\{\mathbf{Z}_t\}$  *icke-singulär* vilket innebär att varje individ inte får exakt ett barn [13]. Då cellerna i populationen delar sig oberoende av varandra enligt en reproduktionsfördelning som endast är beroende av cellens typ, inte tiden, är  $\mathbf{Z}_{t+1}$  endast beroende av  $\mathbf{Z}_t$  och  $\{\mathbf{Z}\}$  är således en Markovkedja. I efterkommande avsnitt introducerar vi allmän teori rörande flertyp-Galton-Watson-processen och tillämpar den för vår specifika modell.

#### 3.1 Väntevärdesmatrisen

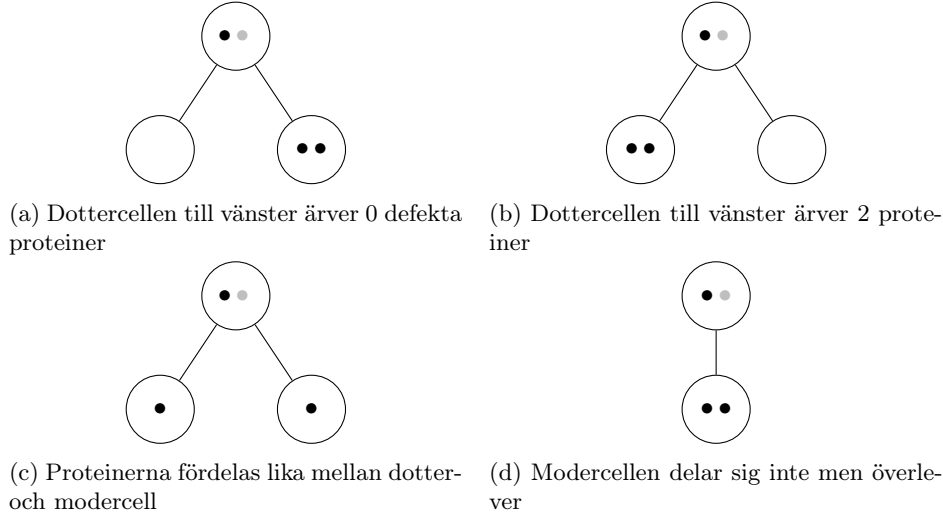
Då cellerna i populationen är av olika typer får cellerna även avkommor av olika typer. Avkommor i detta fall motsvarar antingen moder- och dottercell efter celledelning eller att modercellen inte delar sig men överlever, vilket visualiseras i Figur 3. De möjliga avkommorna beskrivs av *väntevärdesmatrisen*  $\mathbf{M} = (M_{ij})_{i,j=0}^k$  där  $M_{ij} = \mathbb{E}[Z_1^{(j)} \mid \mathbf{Z}_0 = \mathbf{e}_i]$  anger det förväntade antalet celler av typ  $j$  från en cell av typ  $i$  i det första tidssteget. Vidare anger  $(\mathbf{M}^t)_{ij} = \mathbb{E}[Z_t^{(j)} \mid \mathbf{Z}_0 = \mathbf{e}_i]$  det förväntade antalet individer av typ  $j$  efter  $t$  tidssteg givet att processen initiala typ är  $i$  [13].

Vi låter  $i$  vara modercellens typ och  $d = i + n$  vara det totala antalet defekta protein som skall fördelas mellan dottercellerna vid eventuell celledelning. Om modercellen delar sig är  $X \sim \text{Binomial}(d, p)$  där  $X$  anger antalet defekta protein som ärvs från modercellen. Framöver låter vi  $\Pr(X = j)$  ange sannolikheten att dottercellen ärver  $j$  defekta protein från modercellen av typ  $i = d - n$ . I fallet då cellen delar sig kan den få en avkomma av typ  $j \in \{0, 1, \dots, \min(d, k)\}$  på två sätt:

- (a) dottercellen ärver  $j$  defekta protein och modercellen ärver då  $d - j$  defekta proteiner

(b) dottercellen ärver  $d - j$  defekta protein och modercellen ärver då  $j$  defekta proteiner

Notera att om  $d$  är jämnt motsvarar (a) och (b) samma händelse om  $j = d/2$  vilket i sin tur motsvarar två avkommor av typ  $d/2$ . Om moderceller inte delar sig ackumuleras alla  $d$  defekta protein i modercellen vilket motsvarar en avkomma av typ  $d$  om  $d \leq k$ .



Figur 3: Visualisering av de möjliga fallen av avkommor då  $i = 1$ ,  $n = 1$  och  $k = 2$ .

Vi skriver nu väntevärdesmatrisen som  $\mathbf{M} = q\mathbf{A} + \mathbf{B}$  där  $\mathbf{A}$  motsvarar händelserna att cellen delar sig och  $\mathbf{B}$  motsvarar händelsen att cellen inte delar sig. I uttrycket nedan förutsätter vi  $d \leq k$ . Matriselementen ges då av

$$A_{ij} = \Pr(X = j) + \Pr(X = d - j), \quad B_{ij} = \mathbf{1}\{j = d\}(1 - q) \quad (1)$$

Fallet  $M_{ij} = 0$  motsvarar att en cell av typ  $i$  inte kan få en avkomma av typ  $j$  i ett tidssteg. Från definitionen inser vi att om  $(n, k, q)$  hålls fixa är  $\mathbf{M}$  ekvivalent för parametern  $p$  och  $1 - p$ . När vi längre fram i arbetet diskuterar vilket påverkan  $p$  har på populationen är det därför tillräckligt att studera endast halva intervallet  $p \in [0, 1]$ . Vi kommer senare att studera  $p \in (0, 0.5]$  för att det är biologiskt inspirerat att merparten av skadan behålls i modercellen, se avsnitt 2.2. För att överföringen av defekta proteiner skall ske genom en slumpprocess exkluderas  $p = 0$ .

I teorin som presenteras i kommande avsnitt hänvisar vi till egenskaper *positiv reguljäritet* vilket innebär att samtliga typer i populationen kommunicerar. I detta avseende innebär kommunikationen mellan cellerna att det finns en positiv sannolikhet för ättlingar av samtliga typer [13, 3]. I vårt fall gäller det att om cellerna har positiv sannolikhet att dela sig ( $q > 0$ ) så är processen positivt reguljär. En cell av typ  $i$  kan i ett tidssteg få en avkomma av typ  $j \in \{0, 1, \dots, \min(i + n, k)\}$ . Genom ackumulation av defekta protein finns efter  $N$  tidssteg en positiv sannolikhet för en ättling av typ  $j \in \{0, 1, \dots, \min(i + Nn, k)\}$ . Efter  $N = \lceil \frac{n}{k} \rceil$  tidssteg finns således en positiv sannolikhet för ättlingar av samtliga typer och processen är positivt reguljär då samtliga typer kommunicerar.

### 3.2 Garanterad utrotning av populationen

När vi studerar cellpopulationen intresserar vi oss huruvida populationen garanterat kommer att dö ut eller ha en positiv sannolikhet att överleva och växa över tid.

**Sats 3.1** (Perron-Frobenius sats). En strikt positiv matris  $\mathbf{M}$  har ett största egenvärde  $\rho$  som till beloppet är större än alla andra egenvärden och har höger- och vänster egenvektor  $\mathbf{u}$  respektive  $\mathbf{v}$ . Med normaliseringen  $(\mathbf{u} \cdot \mathbf{v}) = 1$  och  $(\mathbf{u} \cdot \mathbf{1}) = 1$  kan man skriva:

$$\mathbf{M}^t = \rho^t \mathbf{P} + \mathbf{R}^t, \quad (2)$$

där elementen i matrisen  $\mathbf{P}$  ges av  $P_{ij} = u_i v_j$  [13, s. 185].



Vi hänvisar genom arbetet till den största positiva egenvärdet  $\rho$  som Perron-Frobenius egenvärdet. Huruvida populationen garanterar kommer att dö ut eller ha en positiv sannolikhet för överlevnad bestäms av storleken på  $\rho$ . Tillståndet  $\mathbf{0}$  är ett absorberande tillstånd och vi säger att populationen utrotats om  $\mathbf{Z}_t = \mathbf{0}$  för något  $t$ . Vi låter  $Q^{(i)} = \Pr(\mathbf{Z}_t = \mathbf{0} \text{ för något } t \mid \mathbf{Z}_0 = \mathbf{e}_i)$  vara utrotningssannolikheten och  $\mathbf{Q} = (Q^{(0)}, \dots, Q^{(k)}) \geq \mathbf{0}$ . Sambandet mellan garanterad utrotning och Perron-Frobenius egenvärdet  $\rho$  ges nu av Sats 3.2 [3]:

**Sats 3.2.** Låt  $\rho$  vara Perron-Frobenius egenvärdet till  $\mathbf{M}$  och att  $\{\mathbf{Z}_t\}$  är positivt reguljär och icke-singulär. Om  $\rho \leq 1$  så gäller det  $\mathbf{Q} = \mathbf{1}$ . Om  $\rho > 1$  gäller  $\mathbf{Q} < \mathbf{1}$ .

Processen kallar subkritisk om  $\rho < 1$ , kritisk om  $\rho = 1$  och superkritisk om  $\rho > 1$ . I det superkritiska fallet växer populationen med en hastighet som bestäms av  $\rho$  [13, s. 186].

### 3.3 Stabila typfördelningen och en populations hälsa

Vidare presenterar vi det asymptotiska beteendet hos cellpopulationer i det superkritiska fallet. Då tiden går mot oändligheten kommer proportionerna av de olika celltyperna i populationen att konvergera mot en stabil typfördelning, som ges av den vänster egenvektor  $\mathbf{v}$  till  $\mathbf{M}$  [3, s. 117].

**Sats 3.3.** Antag att vi har en positiv reguljär flertyp Galton-Watson-process med Perron-Frobenius egenvärde  $\rho > 1$  och att variansen av fördelningen av avkommor är ändlig. Då kommer slumpvektorn  $\mathbf{Z}_t/\rho^t$  konvergera med sannolikhet 1 till en slumpvektor  $\mathbf{W}$ . Om vektorn  $\mathbf{W}$  är nollskild kommer dess riktning med sannolikhet 1 att gå mot  $\mathbf{v}$ , den vänstra egenvektorn av väntevärdesmatrisen  $\mathbf{M}$  [3, s.117].

Fördelningen av typer är intressant att undersöka vid olika parametrar för att se variationen av biologiska åldrar i populationen. I vårt arbete använder vi andelen celler utan defekta proteiner, motsvarande typ-0-celler, som ett mått på hälsan i populationen.

### 3.4 Egenskaper på populations- och individnivå

Med inspiration i tidigare forskning [10] skall vi utifrån vår introducerade modell undersöka hur parameteruppsättningen  $(p, n, k, q)$  påverkar egenskaper hos cellpopulationerna. Vi skall studera den garanterade utrotningen av populationen, populationsstorleken och hälsan i de överlevande populationerna.

Vi betraktar även hur ackumulering av defekta proteiner påverkar åldrandet på individnivå, där åldrande i detta sammanhang avser en cells kronologiska ålder. De individuella cellerna i populationen lever ett slumpmässigt antal tidssteg enligt en livslängdsfördelning. På individnivå får vi en inblick i hur det kronologiska åldrandet påverkas av det biologiska åldrandet. I det stabila fallet gäller  $I \sim \mathbf{v}$  där  $I = i$  anger modercellens initiala typ. Vi kan då beskriva fördelningen av möjliga livslängder för celler i populationen som

$$\Pr(L = \ell) = \sum_{i=0}^k \Pr(L = \ell \mid I = i) \Pr(I = i) = \sum_{i=0}^k \Pr(L = \ell \mid I = i) v_i, \quad (3)$$

där  $L$  anger cellens livslängd. Det är som nämnt är biologiskt inspirerat att merparten av de defekta proteinerna stannar hos modercellen så vi studerar intervallet  $p \in (0, 0.5]$  även på individnivå.

### 3.5 Specialfall av modellen

Vi betraktar även specialfallet  $q = 1$  vilket reducerar parameteruppsättningen hos modellen till  $(p, n, k)$ . Specialfallet motsvarar celler som alltid delar sig i varje diskret tidssteg och är inspirerat av mekanismerna hos jästcellen *S. cerevisiae* vid celldelning. Slumpen i modellen ligger nu endast i fördelningen av defekta protein mellan moder- och dottercell vid celldelning och vi fokuserar på hur asymmetrisk fördelning av protein, motsvarande låga  $p$ -värden i vår modell, påverkar egenskaper på populations- och individnivå. Som nämnt i bakgrunden används ofta replikativt åldrande inom biologi. Eftersom ett tidssteg motsvarar en cellcykel i vår modell så blir kronologisk och replikativt åldrande detsamma i detta fall. Därför diskuteras endast kronologisk ålder.

Med inspiration i tidigare forskning [10] introducerar vi ett föryngringsindex där föryngring definieras som att dottercellens kronologiska ålder är högre än modercellens. Vi definierar  $\varphi_i$  som ett mått den förväntade föryngringsgraden för en dottercell givet modercellens initiala ålder  $I = i$ :

$$\varphi_i = \frac{\mathbb{E}[D - M \mid I = i]}{\mathbb{E}[M \mid I = i]}, \quad (4)$$

där  $D, M$  är dotter- respektive modercellens kronologiska ålder. Vi säger att föryngring sker då  $\varphi_i > 0$ . I det stabila fallet  $I \sim \mathbf{v}$  ges ett föryngringsindex på populationsnivå av

$$\Phi = \mathbb{E}[\varphi_I] = \sum_{i=0}^k \varphi_i v_i. \quad (5)$$

## 4 Metod

Projektet grundar sig i att börja simulera enkla Galton-Watson-processer för att senare utveckla simuleringen och därefter kunna simulera flertyps Galton-Watson-processer med hjälp av programvara. Utökningen från enkel till flertyps Galton-Watson-process bygger på att man får fler parametrar som simuleringen beror på. Denna utökning genomförs med hjälp av den matematiska modellen för åldrande celler som underlag. Valet av programvara anpassas efter kriterier och syfte, till exempel att populationer kan bli stora eller sträcka sig över längre intervall. Det är därför viktigt att tillämpa metoder och använda programspråk som genomför detta på det effektivaste sättet möjligt. Detta är avgörande då Galton-Watson processen växer exponentiellt vid superkritiska tillstånd och blir en krävande process. Risken är annars stor är att simuleringarna tar för mycket plats, därmed mycket tid och kan i vissa fall inte genomföras.

Under arbetets gång används språken R samt C++. Ur ett databehandlingsperspektiv är R lämpligast för enkel och flertyps Galton-Watson-processer eftersom språket har effektiva inbyggda funktioner som är anpassade för statistiska problem, och kan bättre visualisera resultat för detta projekt. Å andra sidan är C++ snabbare för simuleringar då nya objekt inte skapas för varje simulering, vilket leder till reducering av minnesanvändning. En kombination av dessa innebär därför en optimal väg att simulera åldrande celler.

Simuleringar av enkel och flertyps Galton-Watson-processer grundar sig i analytiska metoder för att genomföras. Projektet utgår från ett specialfall av en flertyp Galton-Watson-process där modercellen delar sig med en sannolikhet och får en dotter med en slumpmässig typ. Simuleringarna jämförs med teoretiska resultat för att kontrollera att dessa är korrekta samt för att dra slutsatser. Insamling av den analytiska delen sker även bland möten med handledaren och ämnesexperter för att få en tydligare uppfattning av exempelvis, åldrande celler, defekta protein och Galton-Watson-processer, samt handledning för att utveckla modellen.

I kombination med simuleringar av processen implementeras även de teoretiska resultatet för vänstervärdesmatrisen i R, se avsnitt 3.1. Utifrån detta kan Perron-Frobenius egenvärdet  $\rho$  till  $\mathbf{M}$  och tillhörande vänster egenvektor  $\mathbf{v}$  beräknas numeriskt med hjälp av R:s standardbibliotek.

## 5 Resultat

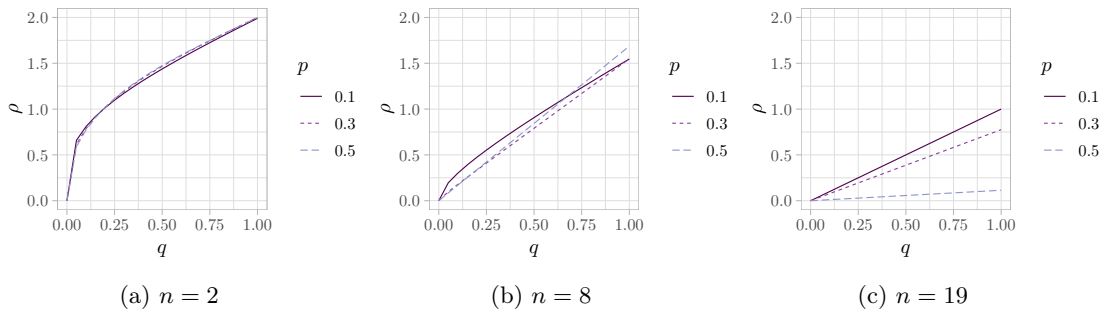
I följande avsnitt presenterar vi de huvudsakliga resultaten som erhållits från numeriska beräkningar och genom simulering av processen. Vi presenterar först hur parameteruppsättningen  $(p, n, k, q)$  påverkar egenskaper hos cellpopulationer och sedan det kronologiska åldrandet hos individuella celler. Slutligen studerar vi föryngring i populationer. Diskussion och tolkning av resultatet genomförs sedan i avsnitt 6.

### 5.1 Kritisk gräns för utrotning

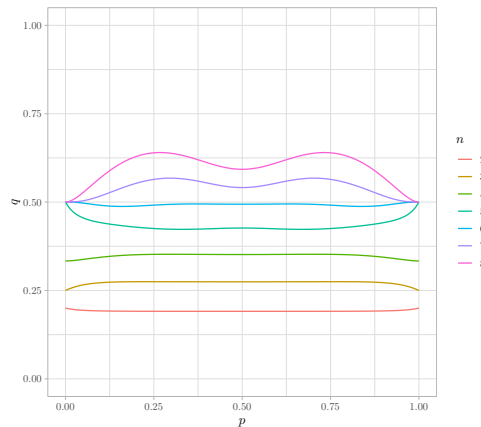
För att studera vilka parameteruppsättningar som garanterar utrotning av populationen tillämpar vi Sats 3.2 som säger att utrotning garanteras för  $\rho \leq 1$ . Det kritiska fallet ges av  $\rho = 1$  och anger

gränsen mellan positiv sannolikhet för överlevnad och garanterad utrotning. Storleken av Perron-Frobenius styrs till stor del av förhållandet mellan  $n$  och  $k$ . I fallet  $n > 2k$  gäller det att  $\rho \leq 1$  oberoende av  $p, q$  och utrotning av populationen är garanterad, se Figur 4 (c). För tillräckligt små  $n$  existerar en kritiskt process, se Figur 4 (a)-(b), för alla  $p \in [0, 1]$  och bestäms av ett tillräckligt litet  $q$ .

Figur 5 visar ett exempel på det  $q$ -värde  $q_\delta$  som ger en kritisk process som en funktion av  $p$  för olika  $n$  på hela intervallet  $p \in [0, 1]$ . Vi beräknar det kritiska  $q$ -värdet på hela intervallet  $p \in [0, 1]$  endast för visualiseringssyfte. Det hade som nämnt varit tillräckligt att endast studera halva intervallet, på grund av symmetri. Givet mer ackumulation av defekta proteiner i varje tidsenhet stiger det kritiska  $q$ -värdet vilket motsvarar att cellerna måste dela sig med större sannolikhet för att undvika garanterad utrotning av populationen.



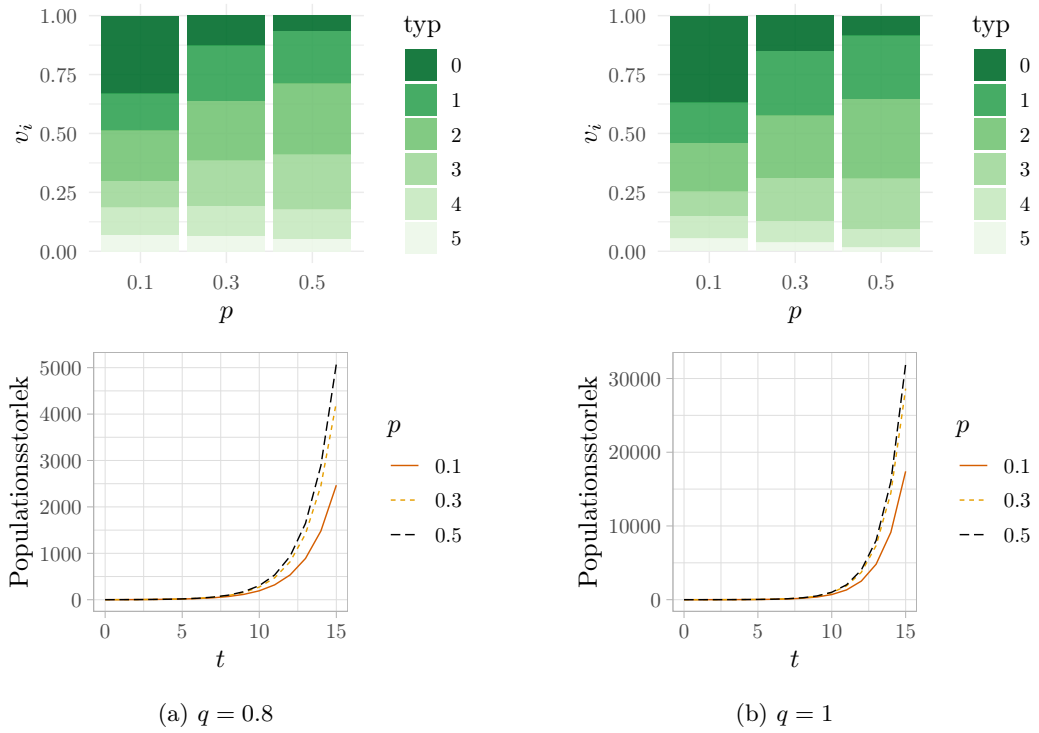
Figur 4: Parameteruppsättningens påverkan på  $\rho$ , då  $k = 9$ .



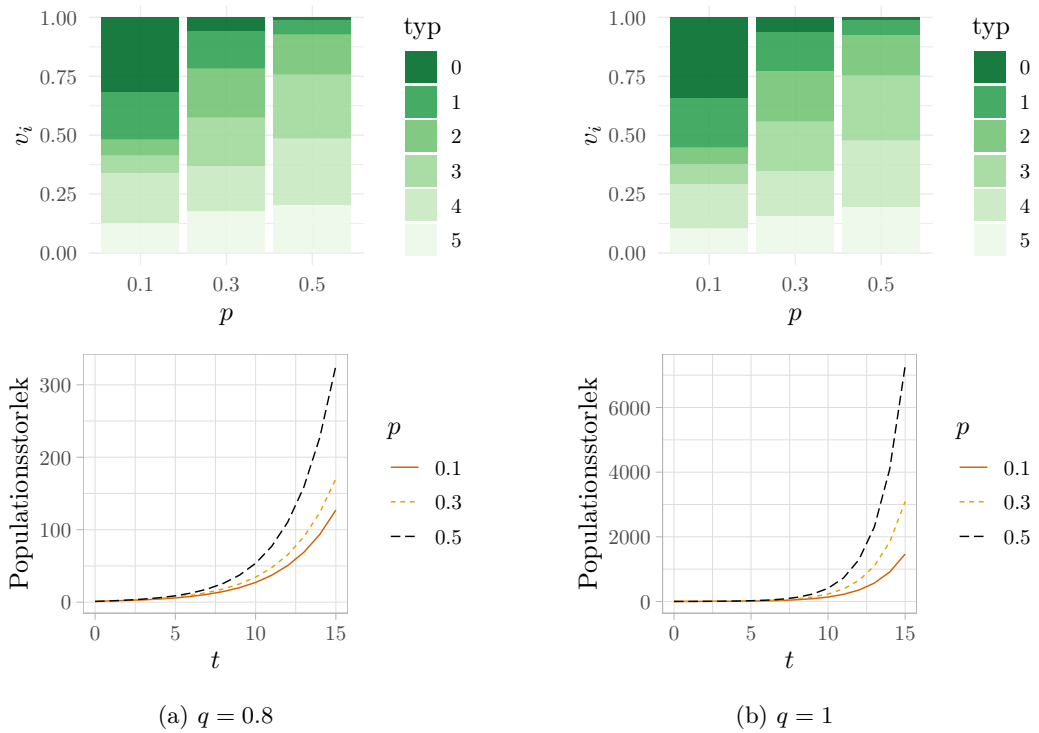
Figur 5: Exempel på det  $q$ -värde som ger en kritisk process ( $\rho = 1$ ) då  $k = 9$ .

## 5.2 Fördelning av biologiska åldrar och populationsstorlek

I Figur 6-7 presenteras hur olika parameteruppsättningar påverkar populationsstorleken samt den stabila typfördelningen  $\mathbf{v}$  i populationen. Vid lägre  $n$ -värden, motsvarande lägre ansamling av defekta protein får vi både större populationer och en högre andelen biologiskt unga celler i populationerna. I våra mått ökar hälsan i dessa populationer. Genom att celler alltid delar sig ( $q = 1$ ) ges även större populationer jämfört med  $q < 1$ . Oberoende av valet av  $q$  och skillnaden mellan  $n$  och  $k$  resulterar en asymmetrisk fördelning av defekta proteiner i populationer celler som har lägre biologiskt ålder. En symmetrisk fördelning av de defekta proteinerna gav emellertid upphov till större populationer. Stora populationer går alltså inte hand i hand med hälsosamma populationer.



Figur 6: Stabil typfördelning och populationsstorlek för olika val av  $p$  och  $q$  då  $n = 2$  och  $k = 5$ .

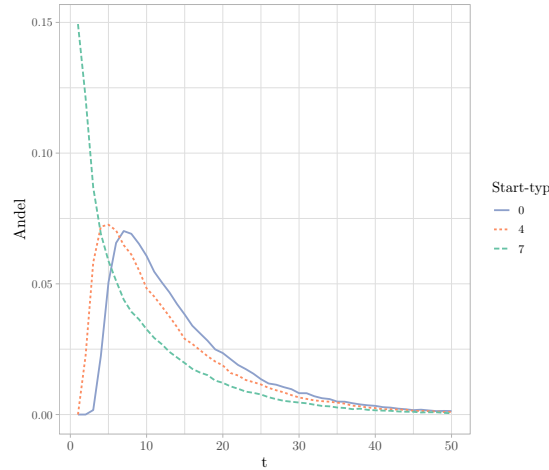


Figur 7: Stabil typfördelning och populationsstorlek för olika val av  $p$  och  $q$  då  $n = 4$  och  $k = 5$ .

### 5.3 Cellers levnadslängdsfördelning

Nedan presenteras resultatet för simuleringen av levnadslängdsfördelningen givet modercellens initiala ålder  $i$  samt för olika parameteruppsättningar. Simuleringen av levnadslängdsfördelningen betingat på modercellens initiala ålder  $I = i$  utgick från simulering av  $10^5$  modercellers levnadslängd. För att finna fördelningen på populationsnivå simulerade vi  $10^5$  moderceller av vardera typ och använde (3) där  $v$  beräknades numeriskt.

Figur 8 presenterar hur cellens initiala biologiska ålder kan påverka cellens livslängd. I grafen är  $x$ -axeln förkortad för att kunna visa grafens form. Ju högre initial biologisk ålder hos cellen desto snabbare dör cellen. Livslängden för alla celler med olika initiala biologiska åldrar ökas vid högre värden på  $p$  men i allmänhet är det förhållandet mellan parametrarna  $n$ ,  $k$ ,  $p$  och modercellens initiala biologiska ålder som gör skillnad i cellernas livslängd.

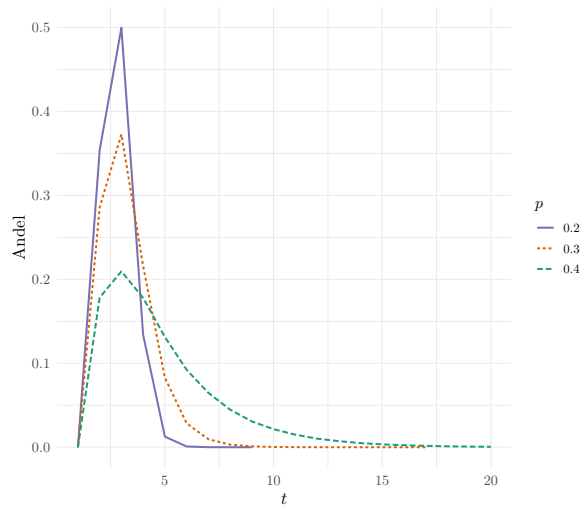


Figur 8: Initial biologisk ålders  $i$  påverkan på levnadslängd. ( $p = 0.3$ ,  $n = 3$ ,  $k = 8$ ,  $q = 1$ ).

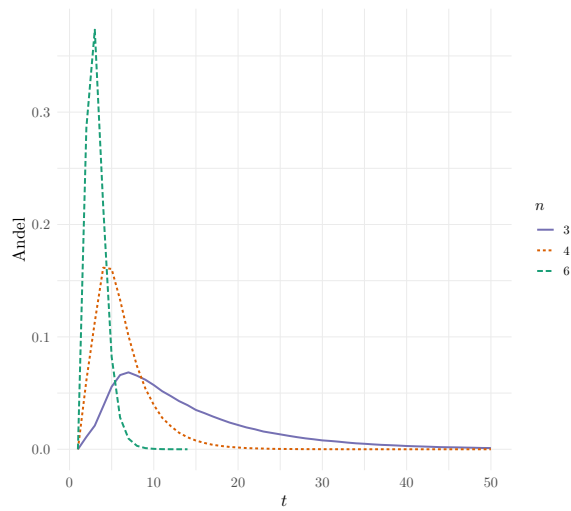
Parameter  $p$  kontrollerar en stor del av livslängden för moder- och dottercellerna. Ju högre värde  $p$  har, desto längre lever modercellerna och majoriteten av de defekta proteinerna ärvs av dottercellerna. Detta ses i figur 9a, där ett högt  $p$ -värde leder till ökade levnadslängder, vilket även innebär fler förekomster av levnadslängder som skiljer sig mycket åt vid små ändringar på parametrarna. Detta leder till att dottercellerna dör mycket fortare än modern. En hälsosam population bedöms utifrån andelen celler av typ 0. Ett lägre  $p$ -värde innebär en kortare medellivslängd för modercellerna, men en högre sannolikhet att fler dotterceller överlever, vilket resulterar i hälsosamma dotterceller. En konsekvens av detta är att förekomsten av orimligt höga åldrar minskar då moderceller inte får chansen att leva tillräckligt länge.

Parametrarna  $n$  och  $k$  har en viktig samverkan som även styr individernas åldersfördelningar. Detta ges genom att sätta  $p$  och  $q$  fixat. Det enklaste fallet är för  $n \leq k$ . Ju större  $k$  är i förhållande till  $n$ , desto längre lever både moder- och dottercellerna. Detta kan ses i fallen för celler med parameteruppsättningarna enligt figur 9b där livslängden är högre för lägre  $n$ . Alltså dör cellerna snabbare ut ju högre  $n$  och lägre  $k$  är.

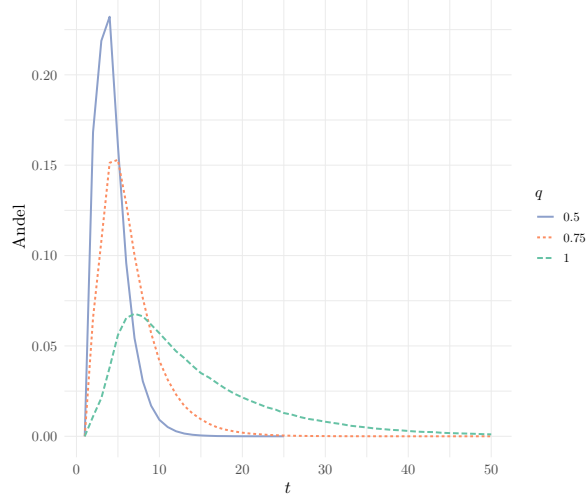
För att moderceller ska kunna fördela defekta protein till sina dotterceller och därmed leva längre krävs det ett högt värde på  $q$ . Enligt figur 7 tyder det på att  $q = 1$  är det optimala fallet där flest celler bildas samt lever längst och är mest lik jästcellen. Konsekvenserna av att  $q < 1$  innebär att antalet cellfördelning reduceras och därmed en levnadslängdsfördelning med en kortare medellivslängd då fler defekta proteiner per cell ackumuleras.



(a) Parametern  $p$ 's påverkan på levnadslängdsfördelningen ( $n = 6, k = 8, q = 1$ )



(b) Påverkan av förhållandet mellan parametrarna  $n$  och  $k$  på levnadslängdsfördelningen ( $p = 0.3, k = 8, q = 1$ )



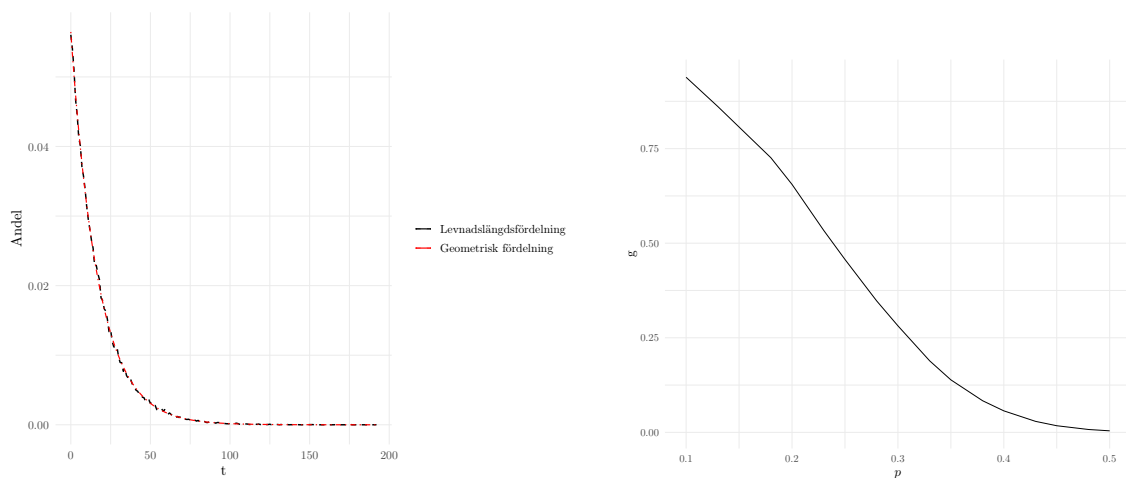
(c) Parametern  $q$ 's påverkan på levnadslängdsfördelningen ( $p = 0.3, n = 3, k = 8$ )

Figur 9: Simulering av levnadslängdsfördelningen för olika parameteruppsättningar

## 5.4 Levnadslängdfördelningens svans följer en geometrisk fördelning

Inspirerat av Holger Rootzén och Dmitrii Zholud artikel tittade vi även på svansen av levnadslängdsfördelningens kurvan för att se om även den följer den geometriska fördelningen. Figurerna nedan presenterar förhållandet mellan cellers livslängd efter ett visst tidssteg och den geometriska fördelningen. Den geometriska fördelningen är en diskret fördelning vilket passar bra med våran modell.

Figur 10a visar att svansen i levnadslängdsfördelningens kurvan följer exakt en geometrisk fördelning med  $g = 0.0567$ . Parametern  $g$  i geometrisk fördelning visar sannolikhet för ett lyckat utfall. I det här fallet kan  $g$  definieras som sannolikheten att en cellen dör efter ett viss tidssteg. Uppskattning av parametern  $g$  görs med maximum likelihood, se avsnitt C.3 i appendix. Figur 10b visar förhållandet mellan sannolikheten  $g$  i den geometriska fördelningen och parametern  $p$ . Ju högre värde på  $p$  desto mindre blir värdet på  $g$ . Värdet på  $g$  är oberoende av cellens initiala biologiska ålder, se tabell 1 i avsnitt A.1 i appendix.



(a) Livslängden typ-0 celler efter 10 tidssteg med parametrar  $p = 0.4$ ,  $n = 4$ ,  $k = 8$ ,  $q = 1$

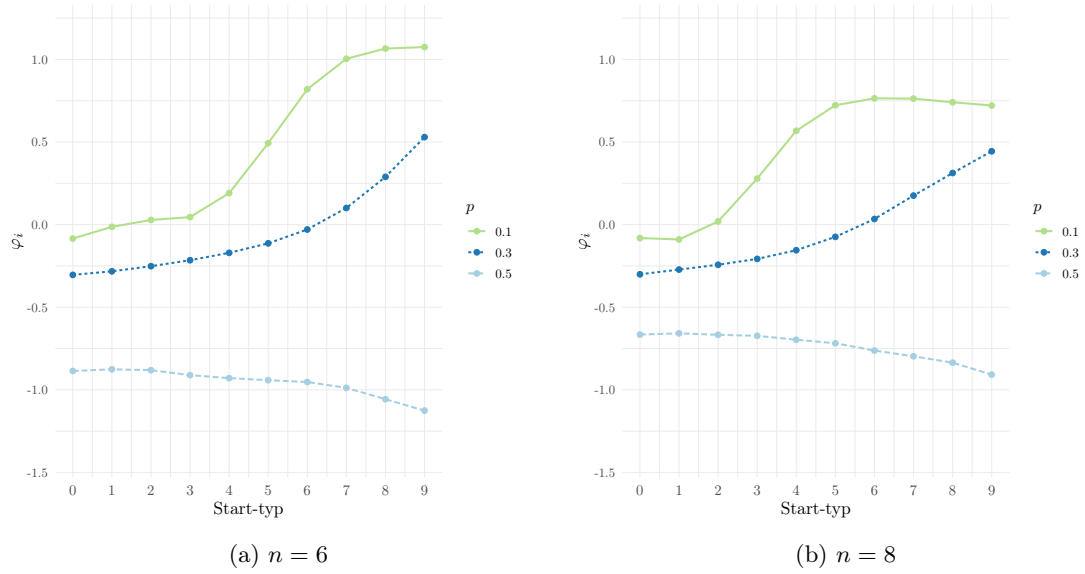
(b) relationen mellan  $g$  och  $p$  med parametrar  $n = 4$ ,  $k = 8$ ,  $q = 1$

Figur 10: Grafer som visar att fördelningen av cellers livslängd följer geometrisk fördelning och relationen mellan värdet på  $g$  och värdet på  $p$

## 5.5 Föryngring inom populationer

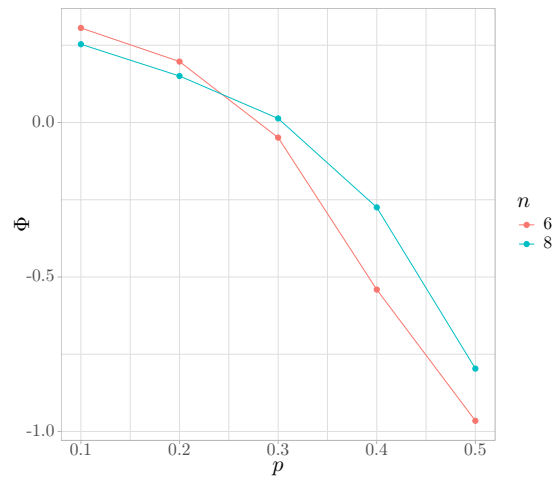
För att uppskatta föryngringsindexet  $\varphi_i$  simulerades  $10^5$  moderceller av varje initial ålder  $i \in \{0, 1, \dots, 9\}$  och därefter simulerades alla dess dotterceller. Även det viktade föryngringsindexet utgick från simulering av  $10^5$  moderceller av vardera typ av viktades sedan enligt (5) där  $v$  beräknas numeriskt.

Figur 11 visar stickprovsmedelvärdet av föryngringsgraden  $\varphi_i$  för en dottercell givet modercellens initiala ålder  $i$ . Vi fann att en asymmetrisk fördelning av de defekta proteinerna ger upphov till en positiv föryngringsgrad vid tillräckligt hög initial biologisk ålder. Detta motsvarar att dotterceller från biologiskt gamla moderceller förväntas bli kronologiskt äldre än modercellen. Vidare kunde en högre grad av föryngring observeras vid lägre  $n$ . Föryngring framträder även i fallet  $p = 0.3$  men endast för biologiskt gamla moderceller. Vid en symmetrisk fördelning av de defekta proteinerna fick vi istället  $\varphi_i < 0$  oberoende av modercellens initiala ålder.



Figur 11: Föryngringsindex  $\varphi_i$  för olika initiala biologiska åldrar då  $k = 9$ .

Figur 12 visar sedan det viktade föryngringsindexet  $\Phi$  som ger en indikation på föryngringen på populationsnivå. Föryngring framträder i populationer där fördelningen sker asymmetriskt mellan moder- och dottercellen. Som förväntat ger en symmetrisk fördelning av defekta proteiner  $\Phi < 0$ .



Figur 12: Viktat föryngringsindex  $\Phi$  för olika val av  $n$  och  $p$  då  $k = 9$ .



## 6 Diskussion

I följande avsnitt diskuterar vi och ger en biologisk tolkning av de resultat som erhållits i avsnitt 5 och kopplar samman de observerade egenskaperna på populations- och individnivå. Vi diskuterar även modellens begränsningar, rimlighet ur ett biologiskt perspektiv och ger förslag på framtida forskning inom området.

### 6.1 Tolkning av resultat

Ett förväntat resultat är att parametern  $q$  blir avgörande för cellpopulationernas överlevnad då det styr huruvida modercellen kan fördela ut skadorna mellan moder- och dottercellen. Vidare minskar storleken på populationen när cellerna inte delar sig tillräckligt frekvent. Vi observerar även att minskade  $q$  leder till en förkortad livslängd hos cellerna i populationen. Det beror främst på att antalet proteiner som ackumuleras i varje tidssteg, som är lika oberoende om modercellen delar sig, måste fördelas till ett mindre antal celler. Över tid ackumuleras de defekta proteinerna och närmar sig gränsen  $k$  och resulterar till slut i celldöd. Fallet  $q = 1$  är alltså gynnsamt i samtliga avseenden och är även biologiskt motiverat hos celler.

Ett lågt  $p$  kan liknas med asymmetrisk fördelning av defekta proteiner och ett högt  $p$  som närmar sig 0.5 liknar en symmetrisk fördelning. Även för lägre  $p$  fås mindre populationer eftersom livslängden minskar för modercellerna på grund av att fler proteiner samlas hos modern än dottercellerna. En annan följd av detta är att populationer blir hälsosammare eftersom andelen typ-0-celler ökar. Något annat som tyder på cellernas hälsa är att föryngring förekommer oftare. I detta fall syftar föryngring på att celler får en ökad kronologisk ålder jämfört med sin moder. I en cellpopulation indikerar detta att nästkommande generationer får mindre skador. Motsatsen gäller för ett högre  $p$  där populationerna blir större men mindre hälsosam. Anledningen bakom detta är att modercellerna lever längre på bekostnad av att överföra en större del av defekta proteiner till sina dotterceller. Populationerna uppnår därför en mindre andel celler av typ-0.

En slutsats som kan dras är att sambandet mellan  $n$  och  $k$  har en stor betydelse för cellers livslängd, då de reglerar ackumuleringen av defekta proteiner. Då värdet på  $k$  ökar, innebär det att celler kan vara av flera olika typer och tolererar fler skador, vilket leder till en längre livslängd. I vissa fall kan dessa livslängder vara orimligt långa. Det är viktigt att notera att dessa orimligt åldrar förekommer oftare ju högre värde  $k$  har eftersom färre begränsningar sätts på cellerna. Däremot är sannolikheten för celler att uppnå dessa åldrar mycket låg. Ett högt värde på  $n$  leder till en ökad ackumulationshastighet av defekta proteiner och orsakar därmed en minskad levnadslängd. På samma sätt fås en ökad levnadslängd för ett lågt  $n$  på grund av den minskade ackumulationshastigheten av proteiner. En följd av relationen mellan  $n$  och  $k$  är att cellerna har den högsta levnadslängderna för låga  $n$  och höga  $k$ . Dessutom ökar andelen typ-0-celler. På samma sätt minskar levnadslängderna för låga värden på  $k$  samt höga värde på  $n$ . Det är därför viktigt att anpassa valen för att uppnå rimliga åldersfördelningar.

Initial biologisk ålder av en cell har stor påverkan på cellens livslängd under de första tidsstegen. Ju högre initial biologisk ålder desto snabbare dör cellen vilket är logisk eftersom en cell som föds gammal och ackumulerar många skador från början inte kan leva länge enligt teorin.

Resultatet visar att svansen av cellernas levnadslängdsfördelningskurva följer en geometrisk fördelning. I den geometriska fördelningen är varje försök oberoende av föregående försök. När svansen av cellernas levnadslängdskurva följer den geometriska fördelningen betyder det att cellernas livslängd efter en viss tid inte är beroende av sina initiala biologiska åldrar och cellerna glömmar sina initiala tillstånd. Man kan koppla detta till teorin om extrema åldrar hos människor. Efter att människor blir äldre än 110 år beror deras livslängd inte längre på deras bakgrund eller på faktorer som kön, genetik eller livsstil.

Parametern  $g$ , som är sannolikheten för ett lyckat utfall i den geometriska fördelningen, minskar när värdet på  $p$  ökar. Den här parametern tolkas som sannolikheten att en modercell dör efter ett visst tidssteg. Man kan dra slutsatsen att när värdet på  $p$  ökar, minskar sannolikheten att cellen dör eftersom cellen överför alla skador och ackumulerade proteiner till sin dotter och blir frisk.

## 6.2 Modellens begränsningar och framtida forskning

Den modell som utvecklats under arbetet ger ett nytt sätt att studera åldrande och ackumulation av defekta protein hos celler med hjälp av sannolikhetsteori. Trots att modellen är såpass enkel och vid vissa resultat till och med reduceras till tre parametrar har man kunnat se vissa resultat som speglas i verkligheten. Vi anmärker dock att modellen fortfarande är för enkel för att dra några slutsatser kring verkliga biologiska system.

Modellen beror på maximalt fyra parametrar så den är väldigt känsligt för parameterintervall. Små skillnader i ackumulation av skador relativt till tolerans för skador eller små skillnader i  $p$  orsakar stora skillnader i ålder. Exempelvis med parametrarna  $p = 0.3$ ,  $n = 4$ ,  $k = 10$ ,  $q = 1$  för 100000 celler som startar med biologisk ålder noll blir medellivslängden 12.45 tidsenheter. Väljer vi  $n = 3$  blir medel-livslängden 49.84 tidsenheter istället. Trots känsligheten går det att finna parameterintervall som ger livslängder rimliga för jästceller.

Den största begränsningen för modellen när det kommer till biologiska resultat är att biologisk ålder endast bestäms från antal defekta protein. Detta innebär att cellernas kronologiska ålder i ett tidssteg inte påverkar dess sannolikhet att dö i nästa tidssteg. Man kan ha ytterligare en parameter som baserat på kronologisk ålder ökar risken för att en cell dör då den blir gammal. Detta skulle kunna lösa en del av känsligheten. En sådan parameter skulle kunna kompensera för alla kända och okända effekter av biologiskt åldrande som inte tas hänsyn till då man endast tittar på defekta protein.

Ett exempel på en biologisk effekt som inte fångas av modellen är förmågan att laga skador i cellen. I verkligheten har organismer förmågan att laga skador [9, s.359]. Här är det enda sättet för att bli av med skador att man ger bort sina skador till en dotter. Detta påverkar säkerligen känsligheten i modellen. Det finns exempel i andra modeller då man har med en förmåga hos celler som kan laga skador [10]. För att tillämpa det i den här modellen skulle varje nyfödd cell kunna ha en förmåga att laga eller ta bort defekta protein som försämras i varje tidssteg. Detta borde i praktiken förlänga livslängden för celler jämfört med våra resultat.

I modellen används andel celler utan defekta protein som mått för hälsa. För att vidare utveckla modellen tror vi att ett mer formellt mått på hälsa som betraktar fler faktorer än bara andel skadefria celler i hänsyn bör användas. Med det mått på hälsa som används här så offrar sig celler för att få friska döttrar. Detta leder i sin tur till en större omsättning av celler. Hur frisk är egentligen en population där levnadslängder offras för en större populations tillväxt?

## 6.3 Avslutning

Det huvudsakliga syftet med arbetet var att studera hur olika typer av åldrande ser ut för cellpopulationer då defekta protein ackumuleras över tid. Till projektets syfte hörde även att studera hur föryngring kan se ut i cellpopulationer. Vi har med hjälp av en ny modell baserad på sannolikhetsteori kunnat studera hur åldrande hos celler påverkas av ackumulering och asymmetrisk fördelning av defekta proteiner. I stora drag uppfylldes projektets syfte.

Den allmänna Galton-Watson-teorin har kunnat tillämpas och vi har funnit en biologisk tillämpning av teorin. Teorin i kombination med simulering av processen har kunnat användas för att beskriva dynamiken hos cellpopulationer och besvara frågeställningarna. Vi kunde se utrotning garanteras vid tillräckligt hög ackumulation av defekta proteiner eller då celler inte delar sig tillräckligt frekvent. I huvudsak har vi observerat att en asymmetrisk fördelning av defekta proteiner mellan moder- och dottercell resulterar i högre andel biologiskt unga celler i populationer. I dessa populationer fann vi även en positiv föryngringsgrad  $\Phi > 0$ , enligt vårt introducerade föryngringsindex. I populationer där föryngring inte skedde var emellertid populationstillväxten större och levnadslängderna hos de individuella cellerna längre. Utöver att endast studera åldrande celler kunde vi även studera fenomen som observerats i forskning rörande det mänskliga åldrandet. Svanarna i levnadslängdsfördelningen följer en geometrisk fördelning och sannolikheten att överleva ytterligare ett tidssteg blir oberoende av bakgrunden.

Avslutningsvis har modellen på ett enkelt men insiktsfullt sätt modellerat åldrandeprocessen hos celler med hjälp av sannolikhetsteori. Åldrandet är en intressant och komplicerad process där hälsa

blir en avvägning mellan individen och populationen. Genom ytterligare utveckling av modellen och justering av dess parametrar hoppas vi kunna beskriva mer komplexa system. Vidare hoppas vi kunna bidra med ökad förståelse kring åldrandeprocessen hos jästcellen *S. cerevisiae* och därigenom åldrandet hos människan.

## Referenser

- [1] Ziegler C. Behl C. *Cell Aging: Molecular Mechanisms and Implications for Disease*. Vol. 1. Springer Berlin, Heidelberg, 2014.
- [2] Bacaër N. *A Short History of Mathematical Population Dynamics*. Vol. 1. Springer, 2011.
- [3] Axelrod D.E Kimmel M. *Branching Processes in biology*. Vol. 19. Springer-Verlag, 2002.
- [4] R. P. Dobrow. *Introduction to Stochastic Processes with R*. John Wiley & Sons, Inc, 2016.
- [5] S. Alsmeyer G. Gröttrup. “Branching within branching I: The extinction problem”. I: (2015).
- [6] Steven N. Austad Edward J. Masoro. *Handbook of the BIOLOGY of Aging*. Vol. 6. Elsevier, 2006.
- [7] Chang H.Y Rando T.A. “Aging, Rejuvenation, and Epigenetic Reprogramming: Resetting the Aging Clock”. I: *Cell* 148.1 (2012), s. 46–57.
- [8] Colloca G m. fl. “Biological and functional biomarkers of aging: definition, characteristics, and how they can impact everyday cancer treatment”. I: *Current Oncology Reports* 22.11 (2020), s. 1–12.
- [9] Alberts B. m. fl. *Molecular biology of the cell*. Vol. 6. Garland Science, 2015.
- [10] Cvijovic M. Schnitzer B. Borgqvist J. “The synergy of damage repair and retention promotes rejuvenation and prolongs healthy lifespans in cell lineages”. I: *PLoS Computational Biology* 16.10 (2020).
- [11] Zholud D. Rootzén H. “Human life is unlimited - but short”. I: *Extremes* 148 (2017), s. 713–728.
- [12] Richard L. Scheaffer Dennis D. Wackerly William Mendenhall III. *Mathematical Statistics with Applications*. Vol. 7. Thomson Learning, 2008.
- [13] P.E. Ney K.B. Athreya. *Branching Processes*. Springer-Verlag, 1972.
- [14] Olofsson P. Andersson M. *Probability, Statistics and Stochastic Processes*. Vol. 2. John Wiley & Sons, Inc, 2012.

# Appendix

## A Övriga resultat

### A.1 Geometrisk fördelning och initial biologisk ålder

Tabell 1: Förhållandet mellan sannolikhet för lyckat utfall i den geometriska fördelningen och initial biologisk ålder ( $p = 0.4$ ,  $n = 4$ ,  $k = 8$ ,  $q = 1$ , 100000 simuleringar).

Initial biologisk ålder	0	2	4	6	8
$g$	0.05664	0.05678	0.05677	0.05668	0.05687

## B Simulering och programkod

### B.1 Programkod

Samtlig kod som skrivit under projektet återfinns på GitHub: : <https://github.com/lottaer/MVEX01-22-01>.

### B.2 Simulering

Den slutgiltiga modellen använde R kombinerat med C++ för simuleringar. Vid vissa prototyper har Python och Matlab använts.

Från början tänkte man hitta moder-dotter par att jämföra genom att simulera en hel population. Ett försök gjordes då varje dotter hade en pekare till sin moder och en egenskap som sa om den var död eller inte. Sedan simulerades en population till ett visst tidssteg. Att beräkna föryngringsindexet på detta sätt blev extremt långsamt och inga fullständiga resultat fick.

Metoden som istället användes var att man följde en moder tills hon dog och sparade typen av alla döttrar som föddes. Sedan simulerar man varje dotter till dess att hon dör och förkastar alla barnbarn till original-cellen. Efter det kan skillnaden i replikativ ålder beräknas mellan modern och alla hennes döttrar. Detta försök upprepas massa gånger. Sedan kan varje moder-dotterpar delas med väntevärdet av moderns replikativa ålder för att skalas.

## C Grundläggande sannolikhetsteori

I följande avsnitt presenterar vi teori och definitioner som har använts i det slutgiltiga arbetet men ansetts tillhöra förkunskaperna. Vi presenterar även en längre uträkning för maximum likelihood uppskattningen för den geometriska fördelningen.

### C.1 Stokastiska processer och Markovkedjor

En process vars utfall inte är deterministiska utan istället är beroende av slumpen kallas för en slumpprocess eller en *stokastisk process*. En stokastisk process är en samling slumpvariabler  $\{X_t, t \in I\}$  där  $I$  är den så kallade indexmängden. I det tidsdiskreta fallet är en stokastisk process en följd av slumpvariabler med tillståndsrum  $\mathcal{S}$  och indexmängden  $I = \{0, 1, \dots\}$  motsvarar diskret tid. Så kallade Markovkedjor är en speciell typ av stokastiska processer där framtida tillstånd endast är beroende av det nuvarande tillståndet, inte det förflutna [4].

**Definition C.1** (Markovkedja). Låt  $X_0, X_1, \dots$  vara en sekvens av diskreta slumpvariabler med tillståndsrum  $\mathcal{S}$  med egenskapen:

$$\Pr(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = \Pr(X_{n+1} = j | X_n = i),$$

för alla  $i, j, i_0, \dots, i_{n-1} \in \mathcal{S}$  och för alla  $n \geq 0$ . Då är sekvensen  $\{X_n\}$  en Markovkedja.

Den minneslösa egenskapen hos Markovkedjor kallas för Markovegenskapen. Övergångarna mellan tillstånd i en Markovkedja beskrivs av den så kallade övergångsmatrisen  $\mathbf{P}$  där

$$p_{ij} = \Pr(X_{n+1} = j \mid X_n = i), \quad i, j \in \mathcal{S},$$

där matrisen trivialt måste uppfylla  $p_{ij} \geq 0 \forall i, j \in \mathcal{S}$  och  $\sum_{j \in \mathcal{S}} p_{ij} = 1 \forall i \in \mathcal{S}$ . Vi diskuterar även kommunikation mellan olika tillstånd som innebär att olika tillstånd går att nå från varandra [14].

**Definition C.2.** Om  $p_{ij}^{(n)} > 0$  för något  $n$  så är tillståndet  $j$  nåbart från tillstånd  $i : i \rightarrow j$ . Om  $i \rightarrow j$  och  $j \rightarrow i$  kommunicerar tillstånden  $i$  och  $j$  och vi skriver  $i \leftrightarrow j$ .

## C.2 Sannolikhetsfördelningar

Nedan presenteras definitioner för sannolikhetsfördelningar som använts under arbetet men ansetts tillhöra förkunskaper.

**Definition C.3.** En slumpvariabel  $X$  är binomialfördelad om  $X$  har sannolikhetsfördelning:

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \mathbb{N}_0,$$

med parametrar  $n \in \mathbb{N}_0$  och  $p \in [0, 1]$  och vi skriver  $X \sim \text{Binomial}(n, p)$ .

**Definition C.4.** Om en slumpvariabel  $X$  har sannolikhetsfördelning:

$$\Pr(X = x) = g(1-g)^x, \quad x \in \mathbb{N}_0, \quad (6)$$

är  $X$  geometriskt fördelad med parameter  $g$  och vi skriver  $X \sim \text{Geom}(g)$ .

I den geometriska fördelningen motsvarar  $X$  antalet misslyckade utfall tills dess att ett lyckat utfall sker och parametern  $g$  anger sannolikheten för ett lyckat utfall.

## C.3 Maximum likelihood för geometrisk fördelning

I detta avsnitt presenteras i korthet beräkningen av maximum likelihood uppskattningen  $\hat{g}$  av parametern  $g$  i den geometriska fördelningen. För att uppskatta värdet av parametern  $g$  kan så kallade *maximum likelihood uppskattning* användas [14].

**Definition C.5.** Låt  $X_1, X_2, \dots, X_n$  vara ett stickprov från en fördelning med sannolikhetsfördelning  $f_g$  då kallas

$$\mathcal{L}(g) = \prod_{i=1}^n f_g(X_i), \quad (7)$$

för *likelihood funktionen*. Värdet  $\hat{g}$  där  $\mathcal{L}$  antar sitt maximala värde kallas för *maximum likelihood uppskattningen* av parametern  $g$ .

För att finna maximum likelihood uppskattningen  $\hat{g}$  av parametern  $g$  i den geometriska fördelningen tecknar vi utifrån (C.4) först likelihood funktionen.

$$\mathcal{L}(g) = \prod_{i=1}^n g(1-g)^{x_i} = g^n (1-g)^{\sum_{i=1}^n x_i}. \quad (8)$$

Vi maximerar istället för likelihood funktionen den så kallade *log-likelihood funktionen*  $\ell(g) = \ln[\mathcal{L}(g)]$  för att underlätta beräkningen. Givet (8) ges log-likelihood funktionen av

$$\ell(g) = \ln[\mathcal{L}(g)] = \sum_{i=1}^n x_i \ln(1-g) + n \ln g. \quad (9)$$

För att finna  $\hat{g}$  deriverar vi log-likelihood funktionen och sätter högerledet lika med 0 enligt standard metoder inom analys vilket ger

$$\frac{d\ell(g)}{dg} = -\frac{\sum_{i=1}^n x_i}{1-g} + \frac{n}{g} = 0 \iff \frac{\sum_{i=1}^n x_i}{1-g} = \frac{n}{g}. \quad (10)$$

Förenkling och utbrytning ger oss

$$\frac{n}{g} = \frac{\sum_{i=1}^n x_i}{1-g} \iff \sum_{i=1}^n x_i = \frac{n}{g} - n \iff \frac{n}{g} = n + \sum_{i=1}^n x_i \implies g = \frac{n}{\sum_{i=1}^n x_i + n}. \quad (11)$$

Maximum likelihood uppskattningen av  $g$  i den geometriska fördelningen ges således av

$$\hat{g} = \frac{n}{\sum_{i=1}^n x_i + n}.$$