

STOKASTISKE MODELLER I POPULATIONSGENETIK

- til kritikken af teoriladede modeller.

Projektrapport:

Lise Odgård Gade

Susanne Hansen

Michael Hviid

Frank Mølgård Olsen

Vejleder:

Jørgen Larsen

IMFUFA
Roskilde Universitetscenter
Postbox 260
4000 Roskilde

STOKASTISKE MODELLER I POPULATIONSGENETIK
- til kritikken af teoriladede modeller.

Projektrapport af:
Lise Odgård Gade, Susanne Hansen, Michael Hviid
og Frank Mølgård Olsen.

Vejleder:
Jørgen Larsen.

IMFUFA tekst nr. 68/83, RUC.
91 sider.
ISSN 0106-6242.

ABSTRACT:

I dette projekt gives en detaljeret gennemgang af to stokastiske modeller fra populationsgenetikken.

Det umiddelbare formål hermed er at give et konkret eksempel, som kan anvendes i debatten om matematiske modellers natur og struktur. Projektet fokuserer specielt på den del af debatten, der drejer sig om teoriladede modeller.

Med udgangspunkt i arbejdsdefinitionen, at teoriladede modeller er modeller konstrueret med en alment accepteret teori som grundlag, arbejder projektet henimod de problemer og muligheder, der er for at udvikle en kritik af teoriladede modeller.

FORORD

Det foreliggende projekt er en analyse af stokastiske modeller i populationsgenetik.

Projektet er en del af matematikuddannelsens 2. faglig-pædagogiske modul, som også kaldes modelopbygningsmodulet.

Vor målgruppe er læsere med interesse for debatten om matematiske modeller, som har et kendskab til sandsynlighedsregning på et niveau svarende til matematikuddannelsens emnekreds i statistik og sandsynlighedsregning. Vi introducerer selv de nødvendige populationsgenetiske begreber, så med en god ordbog er yderligere kendskab til populationsgenetik ikke nødvendig, om end det naturligvis vil være en fordel.

Vi vil gerne takke vor vejleder Jørgen Larsen for grundig og tålmodig vejledning.

Desuden vil vi takke Hanne Østergård fra Landbohøjskolens afdeling for Husdyravl og Genetik for gode råd og kommentarer.

Med venlig hilsen

Lise Odgård Gade
Susanne Hansen
Michael Hviid
Frank Mølgård Olsen

RUC, juni 1983

INDHOLDSFORTEGNELSE

<u>INDLEDNING OG PROBLEMFORMULERING</u>	s. 1	3.4.3 Matematisk generalisation	s. 51
<u>Kap. 1. POPULATIONSGENETIK OG MATEMATISKE MODELLER</u>	s. 4	3.4.4 Generalisationsbegrebet og diffusionsmodellen	s. 52
1.1 Grundlaget	s. 5	<u>Kap. 4. MODELLERNES ANVENDELSE</u>	s. 56
1.2 Grunde til ændringer i genfrekvenser	s. 7	4.1 Modellernes anvendelse i en debat om evolutionsteorien	s. 56
1.3 Hardy-Weinbergs lov	s. 8	4.2 Diffusionsmodellens anvendelse i plante- og dyreavl	s. 61
1.4 Hvad er en matematisk model	s. 10	<u>Kap. 5. DISKUSSION</u>	s. 63
1.5 Dynamiske processer	s. 11	5.1 Modellernes konstruktion	s. 63
1.6 Forskelle mellem deterministiske og stokastiske processer	s. 11	5.2 Modellernes troværdighed	s. 64
1.7 Problemer med stokastiske og deter- ministiske modeller	s. 11	5.3 Modellernes placering i et fagområde	s. 66
1.8 Ændringer i genfrekvenser som stokastisk fænomen	s. 12	5.4 Problemer med modellernes formalside	s. 66
<u>Kap. 2. FORGRENNINGSMODELLEN</u>	s. 13	5.5 Debatten om matematiske modeller	s. 67
2.1 Generel gennemgang af forgrenings- processer	s. 13	<u>Kap. 6. KONKLUSION</u>	s. 69
2.2 Gennemgang af forgreningsmodellen	s. 17	<u>APPENDIX</u> Formidling af stokastiske modeller til gymnasiet og HF	s. 70
2.3 Kritik af forgreningsmodellen	s. 21	Bilag A Udfaldssti for Wienerproces	
2.4 Forgreningsprocessernes historie	s. 26	Bilag B Sandsynlighedens fordeling med tiden	
<u>Kap. 3. DIFFUSIONSMODELLEN</u>	s. 30	Bilag C Middelværdi- og variansberegninger	
3.1 Fragmenter af diffusionsprocessernes historie	s. 30	Bilag D Approximations-analyse	
3.2 Udledning af Kolmogorovs differential- ligninger	s. 32	Bilag E Omskrivning af udtrykket for over- levelsessandsynligheden.	
3.2.1 Wienerprocessen	s. 32	<u>LITTERATURLISTE</u>	s. 90
3.2.2 Generelle tilvækster	s. 38		
3.3 Gennemgang af diffusionsmodellen	s. 41		
3.4 Diffusionsmodellens troværdighed	s. 49		
3.4.1 Generel diskussion af diffusions- modellens troværdighed	s. 49		
3.4.2 Analyse af diffusionsmodellens approximationer	s. 50		

INDLEDNING OG PROBLEMFORMULERING.

Vort valg af modeltype og emneområde skal ses i sammenhæng med den på instituttet løbende debat om matematiske modeller. Det er almindeligt i den modelteoretiske diskussion her på stedet at skelne mellem to typer modeller, nemlig ad hoc-modeller og teoriladede modeller.

Med en ad hoc-model forsøger man ad matematisk vej at beskrive sammenhængen mellem visse forsøgsdata, for så at anvende modellen til at udføre beregninger for områder, hvor man ikke har eller ikke kan have forsøgsdata. Man har her intet særligt kendskab til de mekanismer, der frembringer de pågældende forsøgsdata.

Teoriladede modeller er konstrueret med udgangspunkt i et indgående kendskab til sammenhænge og mekanismer indenfor modellernes genstandsområde. Denne type modeller skabes ikke nødvendigvis for at levere kvantitative forudsigelser som ad hoc-modellerne; men de konstrueres ofte udelukkende for at holde styr på komplekse sammenhænge i den del af virkeligheden, man modellerer. De bliver derfor af stor betydning for teoridannelsen og må siges at bære kvalitative egenskaber. Lad os her præcisere, at teoriladede modeller efter vor definition er modeller, der konstrueres med en verificeret (alment accepteret) teori som grundlag. Yderligere diskussion af disse modeltyper kan f.eks. ses i (11) og (17).

De matematikprojekter, der foreløbigt er gået ind i den modelteoretiske debat, har mest drejet sig om deterministiske ad hoc-modeller.

Med dette projekt forsøger vi at finde frem til, hvordan man som matematiker kan bidrage til en kritisk analyse af teoriladede modeller indenfor et typisk grundforskningsområde.

Vi valgte populationsgenetik af to grunde:

- 1) det er et nyt område for matematikprojekter.
- 2) populationsgenetikken gør brug af stokastiske modeller.

De stokastiske modeller valgtes både af interesse og fordi tidligere projekter allerede har analyseret væsentlige sider ved deterministiske modeller.

For at gøre opgaven mere overkommelig startede vi en emnekreds i statistik og sandsynlighedsregning. Hensigten var, at

gennemføre emnekredsen jvf. bekendtgørelsen via en integration med projektarbejdet.

Vi startede med at opstille tre generelle spørgsmål, som vi ønskede at få besvaret i projektet:

- 1) Findes der stokastiske modeller i populationsgenetik, der er forankret i verificeret teori ?
- 2) I bekræftende fald - hvilken karakter har denne verifikation og denne forankring ?
- 3) Er det muligt at opstille evalueringskriterier for teoriladede modeller ? Eventuelt hvilke ?

Dette kan godt betragtes som det første udkast til en problemformulering.

Efter en almen orientering indenfor emneområdet fandt vi imidlertid ud af, at en behandling af disse spørgsmål ville resultere i et meget biologisk orienteret projekt, hvilket vi ikke var interesserede i.

Konsekvenserne heraf blev, at vi konstruerede en problemformulering, hvormed vi ville kortlægge væsentlige aspekter ved stokastiske modeller. Besvarelsen heraf kunne så eventuelt danne grundlaget for en diskussion af modellernes teoriladethed og være en indikator for, hvorledes modeller af denne type evalueres.

vi opstillede med dette udgangspunkt følgende system af spørgsmål:

- 1) Hvorledes er stokastiske modeller opbygget ?
- 2) Producerer stokastiske modeller troværdig erkendelse indenfor deres genstandsområde ?
- 3) Anvendes en eventuel erkendelse indenfor et til modellerne knyttet fagområde ? Og i bekræftende fald hvordan ?
- 4) Kan der ved konstruktion af stokastiske modeller opstå problemer med modellernes formalside, som fremprovokerer udvikling af ny matematisk erkendelse ?

Disse fire spørgsmål danner rammen om vort projekt og er altså vor problemformulering.

Det er klart, at ved at vælge et passende antal forskellige artede stokastiske modeller og undersøge disse med hensyn til hvert spørgsmål ville man kunne opnå en rimelig generel besvarelse og efter vor mening et rimeligt generelt billede af, hvad

en stokastisk model er. Det ville dog blive et uoverkommeligt stort projekt.

Vi har i stedet valgt at besvare spørgsmålene ud fra en undersøgelse af stokastiske modeller, som behandler et velafgrænset populationsgenetisk problem: Hvad er sandsynligheden for fiksering af mutante gener ?

Dette spørgsmål er underordnet i forhold til problemformuleringen. Det skal udelukkende ses som et middel til at afgrænse modellernes genstandsområde.

Områdets svarhedsgrad taget i betragtning har vi valgt denne afgrænsning, således at vi forhåbentligt kan besvare spørgsmålene rimeligt dybdegående og undersøge hvilke aspekter, der er indeholdt i de enkelte spørgsmål.

Lad os altså pointere, at det ikke er projektets formål at give en besvarelse af problemformuleringen som er gyldig for en vilkårlig stokastisk model. Vi vil derimod besvare spørgsmålene for et par specifikke modeller, for derigennem at anvise en metode til at analysere en vilkårlig stokastisk model.

Projektets opbygning

Kapitel 1 indeholder en introduktion af diverse begreber fra populationsgenetikken samt nogle betragtninger omkring matematiske modeller.

Vi har udvalgt to stokastiske modeller - forgreningsmodellen og diffusionsmodellen - som vi gennemgår i kapitel 2 og 3. Begge kapitler indeholder følgende punkter:

- 1) noget generel sandsynlighedsteori.
- 2) gennemgang af selve modellen.
- 3) kritik af modellen.
- 4) en historisk/matematisk udvikling i tilknytning til modellen.

Dernæst vil vi se på, hvordan modellernes resultater er blevet brugt indenfor populationsgenetikken (kapitel 4).

Ud fra de foregående kapitler vil vi så i kapitel 5 diskutere spørgsmålene i problemformuleringen. Vi vil forsøge at sætte diskussionen i relation til den almindelige modeldebat.

Konklusionen, der vil være en opsummering af kapitel 5, kommer så i kapitel 6.

I appendix A undersøger vi muligheden af at formidle nogle af vore resultater i gymnasiet og HF.

1. POPULATIONSGENETIK OG MATEMATISKE MODELLER

Konstellationen matematik og biologi bliver mere og mere almindelig. I århundreder har fysikken været det dominerende anvendelsesområde for matematikken, men specielt i dette århundrede er der sket en kraftig udvidelse i antallet af anvendelsesområder.

Populationsgenetikken er et af de naturvidenskabelige områder uden for fysikken, hvor matematikken har været i anvendelse længst. Matematikkens indtog i populationsgenetikken fandt sted omkring 1930 og skyldtes især navne som Fisher, Haldane og Wright.

Populationsgenetikken drejer sig om den genetiske sammensætning hos individer og overførslen af gener fra en generation til den næste. Der er tale om gener, der er årsag til let sknelige - lad os kalde dem kvalitative - træk hos mennesker og dyr. Dermed kan man dele individer ind i forskellige typer, f.eks. blå- og brunøjede individer, individer med forskellige blodtyper eller almindeligt farvede individer og albinoer. Dette må ikke forveksles med f.eks. mælkeproduktionen hos køer, vingefang hos fugle eller højde hos mennesker, der varierer kontinuert, endda ofte indenfor et temmeligt stort interval. Studiet af sådanne forhold kaldes kvantitativ genetik.

Som definitionen antyder, drejer populationsgenetikken sig blandt andet om, hvordan genfrekvenser ændrer sig med tiden, og hvilke faktorer der kan ændre dem. Dette betyder, at populationsgenetikken spiller og har spillet en væsentlig rolle indenfor evolutionsteori. De tidligere nævnte populationsgenetikere har skrevet værker, hvis titler illustrerer dette: "The Genetical Theory of Natural Selection" (Fisher, 1930), "The Causes of Evolution" (Haldane, 1932) og "Evolution in Mendelian Populations" (Wright, 1931).

Den meste simple matematiske model i populationsgenetik er nok Hardy-Weinbergs lov, opkaldt efter Hardy og Weinberg, der viste loven uafhængigt af hinanden i 1908. Vi vender tilbage til denne model senere i kapitlet.

Hardy-Weinbergs lov og populationsgenetikken i det hele taget bygger på Mendels love, opkaldt efter den østrigske augustiner-munk Gregor Mendel. Hans resultater blev udgivet i 1866, men

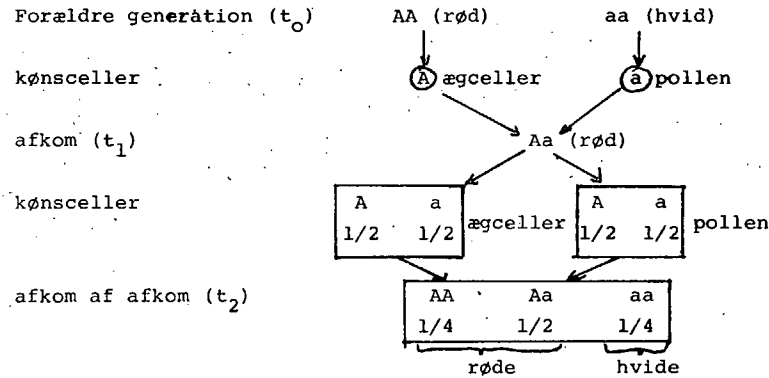
det var først omkring år 1900, man opdagede deres vigtighed.

1.1. Grundlaget

Mendel eksperimenterede med ærteblomster. Som eksempel beskriver vi her hans eksperimenter med blomsterfarver.

Mendel krydsede hvide og røde ærteblomster ved at drysse pollen fra den ene art ud over støvfangene på den anden. Dette resulterede i, at blomsterne i næste generation blev røde. Derefter lod han disse planter selvbefrugte, og det viste sig, at i afkommet var 3/4 af planterne røde, mens 1/4 var hvide.

Fig. 1.1.A. Anskueliggørelse af Mendels fortolkning af sine resultater (3).



Mendel gjorde følgende antagelser:

- 1) Blomsterfarve bestemmes af en faktor (i dag kaldet gen), som kan eksistere i to former (alleler), A og a, der bestemmer henholdsvis røde og hvide blomster.
- 2) Hver plante indeholder to af disse gener - en fra hver af forældrene - således at planten kan have en af tre mulige genotyper: AA, Aa eller aa.
- 3) En kønscelle (ægceller og pollen hos planter eller æg og sæd hos dyr) indeholder kun et gen, som er udtaget tilfældigt fra et genpar (allelpar).
- 4) Ægceller og pollen forenes tilfældigt - uafhængigt af ge-

nerne de bærer - og producerer herved den næste generations planter.

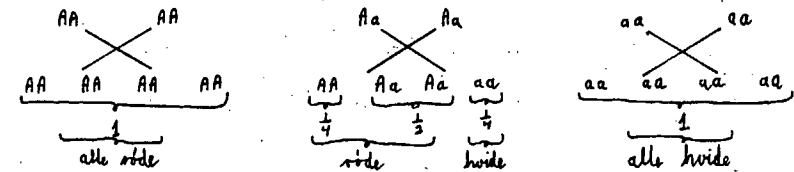
Som det ses af figur 1.1.A. giver Aa-genpar røde blomster. Rød blomsterfarve siges da at være dominant og bestemt af A, mens hvid blomsterfarve siges at være recessiv og bestemt af a.

Det er nyttigt at skelne mellem en plantes genotype (AA, Aa eller aa) og dens fænotype (røde og hvide blomster). Fænotyperne er således det synlige udslag af genotyperne.

Mendel konfirmerede sin teori på to måder:

- 1) Først lod han t_2 -planterne selvbefrugte, så der fremkom en t_3 -generation, hvilket efter teorien skulle foregå efter følgende skema:

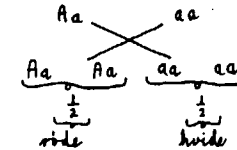
Fig. 1.1.B.



Dette viste sig at stemme !

- 2) Så krydsede han t_1 -planter med hvide blomster i et tilbagekrydsningeksperiment. Dette skulle efter teorien foregå efter følgende skema:

Fig. 1.1.C.



Også her fik han det forventede resultat !

Gennem en række af denne type eksperimenter fik Mendel konfirmeret sin teori.

Som vi har set, var generne hypotetiske konstruktioner, der skulle gøre rede for resultaterne af avlseksemperter. Man blev snart klar over, at der er fysiske objekter - kromosomer

i cellekerner - som opfører sig på nøjagtig samme måde, som det blev postuleret, at generne gjorde. Det blev derfor naturligt at antage, at generne bæres af kromosomerne.

Heraf udvikledes der en omfattende teori, hvorfra stikord som haploid, diploid, locus, homozygot og heterozygot kan nævnes (definitioner: se standardlitteratur). Pointen er imidlertid, at det er cytologiske fakta, som giver en fysisk basis for Mendels love.

1.2. Grunde til ændringer i genfrekvenser

Når generne i en population overføres fra generation til generation kan dens genetiske egenskaber og genfrekvenserne ændres. Dette kan skyldes forskellige ting:

1) Populationens størrelse.

I forældregenerationen "nedbrydes" genotyperne til gener i kønscellerne, hvorefter de sammensættes tilfældigt og danner afkomstgenerationens genotyper. Derfor vil der være en vis tilfældig variation i genfrekvenserne gennem generationer. Jo færre forældre der er, jo større vil den tilfældige variation være. Variationen vil kunne negligeres, hvis populationen er tilstrækkelig stor, iflg Falconer (6) målt i hundreder fremfor i tiere.

2) Forskelligheder i frugtbarhed og overlevelsessevne.

Forskellige genotyper hos forældrene kan have varierende frugtbarhed. Nydannede genotyper kan desuden have forskellige overlevelsessevner, hvilket betyder, at afkomstgenerationen i den kønsmodne alder får en anden sammensætning end forældregenerationen.

Begge disse forhold kan ændre genfrekvenserne og benævnes selektion. Selektion kan både være naturlig og kunstig.

3) Migration og mutation.

Genfrekvenser kan også ændres ved emigration/immigration af individer til/ fra en anden population, samt ved at et gens konstitution ændres - også kaldet mutation.

4) Parringsystemer.

Hvis parring ikke foregår fuldstændig tilfældigt vil genfrekvenserne også ændres. Dvs. at hvis der skal være uændrede genfrekvenser, må visse genotyper ikke have tendens til at fo-

retrække visse andre genotyper, og der må ikke være tendens til, at parrende genotyper er arveligt beslægtede.

Ændringer i genfrekvenserne, der sker på grund af de i punkt 2, 3 og 4 nævnte forhold, kaldes systematiske ændringer i genfrekvenserne, mens ændringer, der sker på grund af populationens størrelse og den tilfældige parring, kaldes tilfældige ændringer i genfrekvensen.

1.3. Hardy-Weinbergs lov

Hvis de i afsnit 1.2. nævnte faktorer, som kan ændre genfrekvenserne i en population, ikke forekommer, kan man formulere en simpel lov for sammenhængen mellem gen- og genotyperfrekvenserne over generationer - den såkaldte Hardy-Weinbergs lov:

I en stor tilfældigt parrende population uden selektion, mutation eller migration er gen- og genotyperfrekvenserne konstante fra generation til generation, og sammenhængen mellem gen- og genotyperfrekvenser er fastlagt ved:

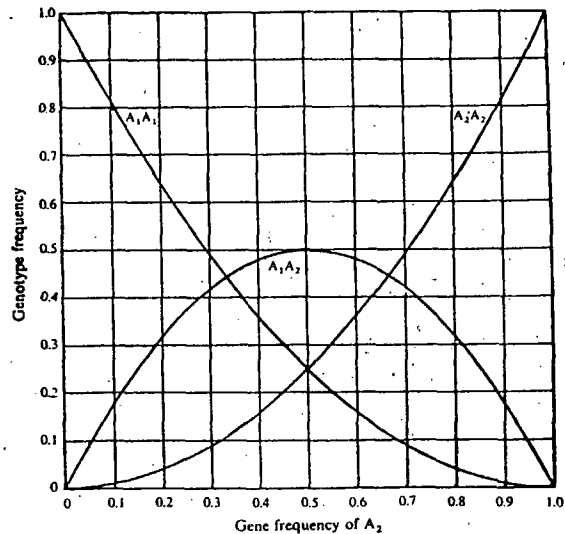
Fig. 1.3.A.

	gener		genotyper		
	A ₁	A ₂	A ₁ A ₁	A ₁ A ₂	A ₂ A ₂
frekvenser:	p	q	p ²	2pq	q ²

hvor $p+q=1$ og $p^2+2pq+q^2=1$.

Hvis vi lader frekvensen af A₂ (q) variere fra 0 til 1 kan frekvenserne for de forskellige genotyper beregnes som funktion heraf, i det p kan bestemmes af q som $p=1-q$. Sammenhængen kan illustreres således:

Fig. 1.3.B.



Fra (6)

Hvis betingelserne for Hardy-Weinbergs lov er opfyldt, vil genotypfrekvenserne fordele sig som loven foreskriver. Fordele genotypfrekvenserne i en population sig efter nedenstående skema

Fig. 1.3.C.

	Alleler		Genotyper		
	A	a	AA	Aa	aa
Frekvenser	120/200	80/200	30/100	60/100	10/100

Indser vi, at Hardy-Weinbergs lov ikke gælder for den pågældende population, da genotypfrekvenserne så skulle have været:

$$p^2 = (120/200)^2 = 9/25; \quad 2pq = 2(120/200)(80/200) = 12/25;$$

$$q^2 = (80/200)^2 = 4/25.$$

Dette skyldes måske, at populationen er for lille til, at tilfældige fluktationer kan ignoreres, men det kan også skyldes nogle af de andre årsager til ændringer i genfrekvenser, der er nævnt i afsnit 1.2.

Med Hardy-Weinbergs lov har vi set den første, simple mate-

matisk model i populationsgenetik. Når vi senere skal se på fikseringssandsynligheder for mutante gener, kan dette opfattes som en undersøgelse af, hvad der sker med genfrekvenserne, når nogle af betingelserne for Hardy-Weinbergs lov svigter.

1.4. Hvad er en matematisk model ?

En matematisk model specificerer en matematisk sammenhæng mellem målte, målelige eller i princippet målelige størrelser (ofte kaldet parametre) i det system eller den proces i virkeligheden, som betragtes.

Modeller laves ikke nødvendigvis for at frembringe kvantitative udtryk for at kunne bestemme et bestemt tal, der skal tjene en teknisk, økonomisk eller lignende interesse, men kan også anvendes til at frembringe kvantitative udtryk for et givet genstandsområde, der kan fortolkes og indgå som en hjælp ved teorikonstruktion. Modeller med disse egenskaber får derved et kvalitativt tilsnit, og matematikkens rolle bliver at være en hjælp for intuitionen, når komplicerede forhold i virkeligheden skal modelleres.

Hardy-Weinberg loven må siges at repræsentere en hel masse erfaring omkring, hvorledes gener opfører sig i en population. Den har basis i Mendels love og cytologiske kendsgerninger, som er verificeret gennem en mangfoldighed af eksperimenter. Man har et præcist kendskab til, hvilke forhold der skal være til stede i en population for at den er i Hardy-Weinberg ligevægt. Man kan derfor trygt sige, at hvis de og de forhold er til stede, så er populationen i Hardy-Weinberg ligevægt. Dette gør det tilladt at slutte, at når populationen ikke er i Hardy-Weinberg ligevægt, så er der noget galt med forholdene, hvilket vi tidligere så et eksempel på. Hardy-Weinberg loven giver således et fingerpeg om, hvornår det er nødvendigt at anvende andre love/modeller til at beskrive udviklingen i gen- og genotypfrekvenser.

Dette præcise kendskab til modellens gyldighedsområde, samt det, at modellen er konstrueret med udgangspunkt i en alment accepteret teori, gør, at vi må karakterisere Hardy-Weinbergs lov som en teoriladet model.

1.5. Dynamiske processer

Det er karakteristisk for levende ting, at de undergår fødsel, vækst, forfald, skift og transformation. De er derfor dybt involveret i dynamiske processer af udvikling i tiden. (1).

Stokastiske processer drejer sig om den dynamiske del af statistisk teori, eller om statistisk forandring i modsætning til, hvad man kunne kalde, statisk statistik. Med en stokastisk proces mener man en mulig virkelig, f.eks. fysisk, proces i den virkelige verden, der har et tilfældigt eller stokastisk element involveret i sin struktur. (2, s.1).

1.6. Forskelle mellem deterministiske og stokastiske processer

Vil man behandle populationsvækst deterministisk, antager man ofte, at væksten er kontinuert både med hensyn til populationsstørrelse og tiden. Man antager også, at tilfældige statistiske fluktuationer kan ignoreres. Sådanne antagelser er ofte rimelige, hvis populationerne er tilstrækkeligt store, og selvom der er tale om betragtelige fluktuationer, kan man ofte anvende deterministiske ligninger ved at betragte dem som beskrivende gennemsnitsværdier.

Vil man derimod lave et kvantitativt studium af 10 avlende par af en sjælden fugleart, er det klart, at tilfældighed vil spille en hovedrolle med hensyn til faktorer som frugtbarhed, kuld størrelse, fødeforsyning etc. Enhver beskrivelse af mulige ændringer i disse forhold må tage hensyn til, at tingene kan udvikle sig ret forskelligt. Derfor må vi til ethvert tidspunkt betragte en sandsynlighedsfordeling af populations størrelse, og der er da tale om en stokastisk proces i modsætning til en deterministisk, som til et givet tidspunkt knytter en bestemt (gennemsnits-)værdi.

1.7. Problemer med stokastiske og deterministiske modeller

Selv en simpel, stokastisk model kan give anledning til en mængde matematisk komplikation. Den simple model giver dog ikke altid tilstrækkelig information om virkeligheden, så derfor forsøger man ofte at gøre modellen mere realistisk,

hvilket giver anledning til endnu flere matematiske komplikationer. Man kan løbe ind i et dilemma, hvor en model, der kan behandles udtømmende matematisk, ikke giver tilstrækkelig information om virkeligheden, mens en model, som kan beskrive genstandsområdet udtømmende, ikke kan behandles matematisk.

Deterministiske modeller har nøjagtig det samme problem, men en deterministisk model vil i alle tilfælde være simplere end dens stokastiske analog.

1.8. Ændringer i genfrekvenser som stokastisk fænomen

Som nævnt er vi interesserede i stokastiske modellers natur og struktur. Først må det præciseres, hvilke faktorer i genstandsområdet, der nødvendiggør en stokastisk behandling.

Man kan sige, at der specielt er to former for årsager til ændringer i genfrekvenser de matematiske modeller kan behandle (4) :

- 1) Konstante mutations- og migrationsrater samt konstant selektionspres resulterer direkte i ændringer. Sådanne systematiske pres på genfrekvenser kan isoleret set behandles deterministisk.
- 2) Tilfældige fluktuationer i genfrekvenser, som skyldes tilfældig forening af kønsceller i en mindre (endelig) population. Selv i en meget stor (uendelig) population afhænger mutante geners skæbne meget af tilfældigheder, så længe disse gener optræder i et ringe antal.

En behandling af disse forhold må nødvendigvis være stokastisk.

De stokastiske modeller, vi ser på i det følgende, beskriver genfrekvensernes udvikling i tiden.

Den første - forgreningsmodellen - behandler den tilfældige fluktuation i frekvenser af mutante gener i en meget stor population, hvor der er konstant selektionspres gennem tiden. Den anden - diffusionsmodellen - behandler det tilsvarende problem i en vilkårligt stor population.

2. FORGRENINGSMODELLEN.

Ved brug af forgreningsprocesser vil vi i dette kapitel udføre beregningerne til fastlæggelse af $u(s)$ - sandsynligheden for fiksering af én mutation som funktion af selektionskoefficienten. Dette blev første gang gjort af henholdsvis Haldane i 1927 og Fisher i 1930. Vi vil opbygge kapitlet således, at vi i afsnit 2.1 udleder et udtryk for udslettelsesandsynligheden vha. en generel teori for forgreningsprocesser. Sidenhen i afsnit 2.2 vil vi anvende dette udtryk til at få et for genstandsområdet specifik udtryk for overlevelsessandsynligheden, hvor overlevelsessandsynligheden er 1 minus udslettelsesandsynligheden. Efter vi har udledt udtrykket for $u(s)$ vil vi i afsnit 2.2 give en kritik af forgreningsmodellen, og i sidste afsnit i dette kapitel vil vi sætte udledningen af modellen i relation til udviklingen af den generelle teori for forgreningsprocesser.

2.1 Generel gennemgang af forgreningsprocesser.

Forgreningsprocesser er en klasse af Markov-kæder med diskret tid og tilstandsrum, hvor en Markov-kæde er defineret ved

$$P\{X_{t+1} = j \mid X_t = x_t, \dots, X_0 = x_0\} = P\{X_{t+1} = j \mid X_t = x_t\}$$

dvs at enhver tilstand X_{t+1} kun er afhængig af den sidst kendte tilstand.

I en forgreningsproces kan man opfatte de stokastiske variable $\{X_0, X_1, \dots, X_t, \dots\}$ som en populationsstørrelse til et bestemt tidspunkt $\{0, 1, \dots, t, \dots\}$. Hvert enkelt individ, som optræder i forgreningsprocessen er i stand til at producere et tælleligt men ikke nødvendigvis begrænset antal efterkommere af samme art, og processen kan således udvikle sig til det uendelige med tiden. Desuden er hvert enkelt individs reproduktionsevne uafhængig af de andre individer.

Vi betragter en population, som nedstammer fra et enkelt individ til tiden $t=0$, og er interesseret i at undersøge dette individs mulighed for at udvikle sig til en evig overlevende population. Den stokastiske variabel X_t lader vi betegne antallet af individer i generation t . Vi lader endvidere $p_0, p_1,$

p_2, \dots være sandsynligheden for, at det første individ har formeret sig til henholdsvis 0, 1, ... individer. Vi kan skrive

$$p_k = P(X_1 = k \mid X_0 = 1)$$

v_t er sandsynligheden for populationen er udslettet senest i generation t , og vi kan skrive

$$v_t = P(X_t = 0 \mid X_0 = 1)$$

Hvoraf følger, at $p_0 = v_1$

Vi er interesseret i forløbet af v_t og specielt om der er en grænseværdi

$$V_\infty = \lim_{t \rightarrow \infty} v_t$$

der betegner populationens endelige skæbne.

Vi vil først opskrive en rekursionsformel for v_t , dvs. v_{t+1} som funktion af v_t . Som begyndelse har vi

$$v_{t+1} = P(X_{t+1} = 0 \mid X_0 = 1)$$

som vi regner lidt videre på

$$\begin{aligned} P(X_{t+1} = 0 \mid X_0 = 1) &= \sum_{k=0}^{\infty} P(X_{t+1} = 0, X_t = k \mid X_0 = 1) \\ &= \sum_{k=0}^{\infty} \frac{P(X_{t+1} = 0, X_t = k, X_0 = 1)}{P(X_0 = 1)} \\ &= \sum_{k=0}^{\infty} \frac{P(X_{t+1} = 0 \mid X_t = k, X_0 = 1) P(X_t = k, X_0 = 1)}{P(X_0 = 1)} \\ &= \sum_{k=0}^{\infty} P(X_{t+1} = 0 \mid X_t = k, X_0 = 1) P(X_t = k \mid X_0 = 1) \end{aligned}$$

Vi ser, at

$$P(X_t = k \mid X_0 = 1) = p_k$$

Yderligere har vi

$$\begin{aligned} P(X_{t+1} = 0 \mid X_t = k, X_0 = 1) &= P(X_{t+1} = 0 \mid X_t = k) \\ &= P(X_t = 0 \mid X_0 = k) \end{aligned}$$

idet vi forudsætter, at processen er tidshomogen. Da det yderligere gælder, at hvert enkelt individs skæbne er uafhængig af antallet af individer og deres skæbne, har vi

$$P(X_t = 0 | X_0 = k) = V_t^k$$

og dermed har vi en rekursionsformel for v_t , idet vi kan skrive

$$V_{t+1} = \sum_{k=0}^{\infty} P_k V_t^k = f(V_t)$$

Vi har altså en funktion f , der udtaler sig om udslettelses-sandsynligheden i generation $t+1$ givet udslettelsessandsynligheden i generation t .

I det øjeblik f er kendt kan vi regne sandsynlighederne ud for, at et individ har formeret sig til 0, 1, ... individer i næste generation; dvs. funktionen frembringer sandsynligheder og hedder meget apropos den sandsynlighedsfrembringende funktion for afkomstfordelingen p_0, p_1, \dots . Vi finder afkomstfordelingen ved at differentiere 1, 2, 3, ... osv. k gange og sætte $v_t = 0$. En nærmere argumentation for denne fremgangsmetode kommer i afsnit 2.2, og vi vil kun kort her vise princippet. Vi har

$$f(V_t) = P_0 + P_1 V_t + P_2 V_t^2 + \dots + P_k V_t^k + \dots$$

til udregning af p_0 differentierer vi 0 gange og sætter $v_t = 0$ og får

$$P_0 = f(0)$$

og videre

$$P_1 = f'(0)$$

$$P_k = f^{(k)}(0) / k!$$

For at udregne p_0, p_1, \dots, p_k kræver det, at man kender funktionen f . Oftest er det omvendt, hvor man fastsætter en afkomstfordeling og derudfra får et kendskab til funktionen.

Ved at kigge på forløbet af denne funktion i intervallet $[0, 1]$ kan vi få nogle interessante oplysninger om v_{∞} .

Vi ser først, at $f(0) = p_0$ og $f(1) = 1$

Hvis $p_0 = 0$ er $v_{\infty} = 0$ (individet overlever som population).

Hvis $p_0 = 1$ er $v_{\infty} = 1$ (individet uddør efter første generation)

Vi kigger på intervallet $0 < p_0 < 1$

Vi differentierer først f henholdsvis en og to gange

$$f'(V_t) = \sum_{k=0}^{\infty} P_k k V_t^{k-1}$$

$$f''(V_t) = \sum_{k=0}^{\infty} P_k k(k-1) V_t^{k-2}$$

Da $f'(v_t) > 0$ er f voksende og

da $f''(v_t) > 0$ er f' voksende og dvs., at f er konveks.

Det ses yderligere, at

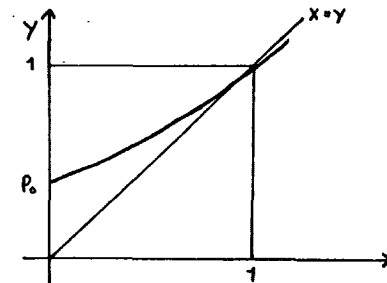
$$f'(1) = \sum_{k=0}^{\infty} k P_k$$

hvilket er middelværdien i afkomstfordelingen p_k , altså det

tal, der angiver middel-antal afkom per individ. Dette tal

kalder vi M . Der er to muligheder for kurvens forløb for funktionen f afhængig af om $M \leq 1$ eller $M > 1$

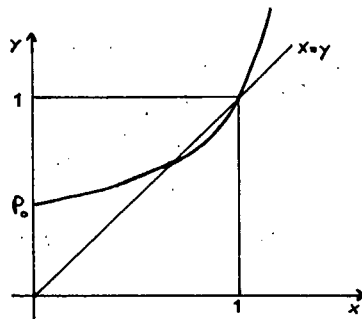
Fig 2.1.A. $M \leq 1$



Er $M = f'(1) \leq 1$ vil kurven ikke skære linjen $x = y$ før punktet $(1, 1)$. (Havde den gjort det, måtte $f'(1)$ nødvendigvis være større end 1, da linjen $x = y$ har hældningen 1). Vi ser, at f er en kontraktion af $[0, 1]$ med fixpunkt $(1, 1)$. Er $f'(1) \leq 1$ har vi

$$\lim_{t \rightarrow \infty} V_t = v_{\infty} = 1$$

Populationen vil med sandsynligheden 1 blive udslettet, såfremt dens middel-antal afkom er mindre end eller lig med 1.

Fig. 2.1.B. $M > 1$ 

Er $f'(1) = M > 1$ må kurven have skåret linjen $x = y$ inden punktet $(1,1)$. I dette tilfælde er f også en kontraktion, dog kun af $[0,1[$ og denne gang med fixpunkt som det første skæringspunkt med $x = y$.

Vi har nu vha. forgreningsprocesser fået udslettelsessandsynligheden udtrykt ved den sandsynlighedsfrembringende funktion, og ved at studere denne funktion i intervallet $[0,1]$ har vi fundet frem til, at for $M \leq 1$ vil individet som population have udslettelsessandsynligheden 1, og for $M > 1$ vil den altid have en positiv sandsynlighed for overlevelse.

2.2 Gennemgang af forgreningsmodellen.

I dette afsnit vil vi gennemgå en model, hvormed en mutants evige overlevelsessandsynlighed bestemmes. Vi anvender den matematiske teori fra afsnit 2.1., idet vi nu lader X_t betegne antallet af mutante gener på et bestemt locus i generation t .

For at få nogle mere præcise oplysninger for udslettelsessandsynligheden V_t for $M > 1$, må man benytte en fordelingstype for afkomstfordelingen p_k , dvs. man må foretage et valg, som er betinget af nogle forhold i genstandsområdet. Haldane (8) og Fisher (7) antager, at afkomstfordelingen for hvert enkelt mutant gen er en poissonfordeling.

$$p_k = \frac{M^k}{k!} e^{-M}$$

Dette udtryk indsætter vi i funktionsforskriften for f og får

$$f(V_t) = \sum_{k=0}^{\infty} \frac{M^k}{k!} e^{-M} V_t^k = e^{-M} \sum_{k=0}^{\infty} \frac{(M V_t)^k}{k!}$$

$$= e^{-M} e^{M V_t} = e^{M(V_t - 1)}$$

Vi indfører u_t som sandsynligheden for overlevelse til tiden t , og der vil da gælde

$$u_t = 1 - V_t \Leftrightarrow V_t = 1 - u_t$$

som indsættes i udtrykket for f og vi får

$$f(V_t) = e^{-u_t M}$$

Da sandsynligheden for udslettelse v_{∞} tilfredsstill

$$f(v_{\infty}) = v_{\infty}$$

tilfredsstill sandsynligheden u for evig overlevelse ligningen

$$2.2.C. \quad 1 - u = e^{-uM}$$

vi lader selektionkoefficienten s være givet ved

$$2.2.D \quad s = \ln M \Leftrightarrow M = e^s$$

$s > 0$ svarer således til en selektiv fordelagtig mutation, $s = 0$ svarer til en selektiv neutral mutation og $s < 0$ til en selektiv ufordelagtig mutation, hvilket betyder, at $u = 0$.

Fodnote: Antagelsen, $s = \ln M$ kan godtgøres på følgende måde

$$E X_t = M^t$$

$$s = \frac{d}{dt} \ln E X_t = \frac{d}{dt} \ln M^t = \ln M$$

som i det diskrete tilfælde bliver

$$s = \ln E X_{t+1} - \ln E X_t = \ln M$$

Udfra ligning 2.2.C og 2.2.D kan vi få u udtrykt som funktion af s. Vi skriver u(s) op som en uendelig række, men inden er vi dog nødt til at opstille to forudsætninger for funktionen u(s)

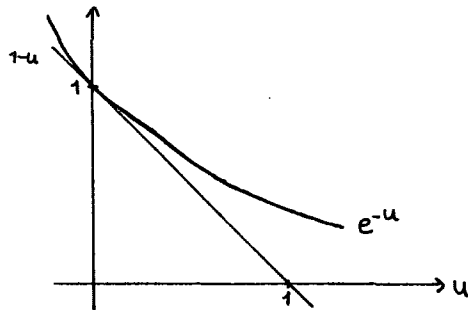
- 1) u(s) skal være differentiabel vilkårlig mange gange
- 2) u(s) skal konvergere mod en fast værdi.

$$2.2.F \quad U(s) = a_0 + a_1s + a_2s^2 + a_3s^3 + \dots = \sum_{k=0}^{\infty} a_k s^k$$

Vores problem bliver nu at bestemme koefficienterne i rækken, og vi kan passende starte med at bestemme a_0 . Hvis $s = 0$ ses af 2.2.F, at $a_0 = u(0)$, men det ses også af 2.2.D, at $M = 1$, som ved indsættelse i 2.2.C giver følgende ligning, hvis løsning er $u(0)$

$$1 - u = e^{-u}$$

Fig 2.2.G.



Tegningen tyder på $u = 0$ er den eneste løsning til ligningen, og dette kan vi nu vise. Først indser vi, at

$$1 - 0 = e^{-0}$$

er sand. Dernæst differentierer vi e^{-u} henholdsvis en og to gange

$$\frac{\partial e^{-u}}{\partial u} = -e^{-u}$$

$$\frac{\partial^2 e^{-u}}{\partial u^2} = \frac{\partial (-e^{-u})}{\partial u} = e^{-u}$$

Heraf ser vi, at e^{-u} har hældningen -1 i punktet $u = 0$, og da den dobbelte afledede altid vil være positiv, og funktio-

nen for e^{-u} dermed er konveks, vil linjen $1 - u$ derfor være tangent til funktionen for e^{-u} , og punktet $(0,1)$ vil være deres eneste skæringspunkt. Vi har altså $a_0 = 0$.

Ved udregningerne af a_1, a_2, \dots bruger vi en generel regel for vilkårlig ofte differentiable funktioner

$$u(0+s) = \sum_{n=0}^{\infty} \frac{u^{(n)}(0)}{n!} \cdot s^n$$

Vi ser ved at sammenligne med 2.2.F, at a_n udspalter sig som

$$\frac{1}{n!} u^{(n)}(0)$$

Vi skal nu bruge en metode, der kan give os henholdsvis den første afledede, anden afledede osv. og dermed a_1, a_2, \dots . Vi starter med at isolere M i 2.2.C og får

$$M = \frac{-\ln(1-u)}{u}$$

hvilket vi kan rækkeudvikle til

$$M = 1 + \frac{u}{2} + \frac{u^2}{3} + \frac{u^3}{4} + \dots$$

Dette indsættes i ligningen $M = e^s$, og vi differentierer på begge sider af lighedstegnet

$$2.2.H. \quad e^s \cdot 0 + \frac{1}{1} \frac{\partial u}{\partial s} + \frac{2}{3} \frac{\partial u}{\partial s} u + \frac{3}{4} \frac{\partial u}{\partial s} u^2 + \dots$$

Lader vi nu $s = 0$ ved vi fra udregningen af a_0 , at $u(0) = 0$, hvilket vi indsætter i 2.2.H

$$e^0 = \frac{1}{1} \frac{\partial u}{\partial s}(0) \Leftrightarrow \frac{\partial u}{\partial s}(0) = 2$$

dvs.

$$a_1 = \frac{2}{1!} = 2$$

Vi gentager succesen og differentierer 2.2.H en gang til

$$e^s = \frac{1}{2} \frac{\partial^2 u}{\partial s^2} + \frac{2}{3} \left(\left(\frac{\partial u}{\partial s} \right)^2 + \frac{\partial^2 u}{\partial s^2} u \right) + \frac{3}{4} (\dots) + \dots$$

Vi indsætter igen $s = 0, u = 0$

$$1 = \frac{1}{2} \frac{\partial^2 u}{\partial s^2}(0) + \frac{2}{3} \left(\left(\frac{\partial u}{\partial s}(0) \right)^2 \right) \Leftrightarrow$$

$$1 = \frac{1}{2} \frac{\partial^2 u}{\partial s^2}(0) + \frac{2}{3} \cdot 4 \Leftrightarrow$$

$$\frac{\partial^2 u}{\partial s^2}(0) = -\frac{10}{3}$$

Hvilket giver

$$a_2 = \frac{1}{2!} \cdot (-\frac{10}{3}) = -\frac{5}{3}$$

På tilsvarende måde kan a_3, a_4 osv. udregnes

Med værdierne for de 4 første koefficienter indsat i 2.2.F får vi (4)

$$U(s) = 2s - \frac{5}{3}s^2 + \frac{7}{4}s^3 - \frac{121}{540}s^4 + \dots$$

Det ses, at når s er større end nul og lille er en god approximation

$$U(s) \approx 2s$$

2.3 Kritik af forgreningsmodellen.

I dette afsnit vil vi analysere de forbehold, der har været nødvendige at lave for at udlede udtrykket for $u(s)$ ved hjælp af forgreningsmodellen. Vi vil først analysere de forudsætninger, som knytter sig til selve forgreningsprocessen, og derfor vil være gældende for en hvilken som helst model, der bygger på forgreningsprocesser. Sidenhen vil vi kritisere selve konstruktionen af modellen og opstille de betingelser, som skal være opfyldt for, at modellen bliver god, eller sagt på en anden måde, er i stand til at give et brugbart resultat. I kritikken af modellen vil vi henvise til en artikel skrevet af Kojima og Kelleher i 1962 (16), hvor de også udfra forgreningsprocesser konstruerer en model til udledning af et udtryk for overlevelsessandsynligheden for mutante gener. De benytter blot en anden afkomstfordeling, nemlig den negative binomialfordeling, istedet for poissonfordelingen.

Til selve forgreningsprocessen bliver der lavet 3 forbehold, som alle vil være gældende for enhver forgreningsproces.

- 1) Populationen skal kunne udvikle sig til det uendelige
- 2) Mutanternes udvikling skal opfylde Markov-egenskaben
- 3) Hver mutants skæbne skal være uafhængig af de andres.

Det første forbehold giver sig selv, nemlig at antallet af mutanter kan udvikle sig til det uendelige. Det andet forbehold betyder, at processen er uden hukommelse, og således kun afhænger af tilstanden før den nuværende. Det sidste forbehold betyder, at der ikke må være konkurrence mellem mutanterne, og der ikke må være nogen ydre væksthæmmende faktor.

For at få udslettelsessandsynligheden udtrykt ved en rekursionsformel, laver vi endnu et forbehold.

Processen skal være tidshomogen.

Dette forbehold forudsætter, at mutantens ydre miljø skal være konstant, hvilket vil sige, at mutantens overlevelsesmuligheder skal være de samme til ethvert tidspunkt. Dette er ensbetydende med, at der ikke må finde nogen former for migration sted, at parringen skal være tilfældig, og at der ikke må ske flere mutationer, - altså en stor indsnævring af modellens anvendelsesområde.

Senere i modellen tager vi endnu et forbehold, nemlig at afkomstfordelingen for de mutante gener er en poissonfordeling

$$P_k = \frac{M^k}{k!} e^{-M} \quad (k = 0, 1, 2, \dots)$$

Dette gør vi for at få et for genstandsområdet specifikt udtryk for udslettelsessandsynligheden og dermed overlevelsessandsynligheden. Udtrykket for p_k indsættes i funktionsforskriften for f , som er en rekursionsformel for v_t , og vi bliver dermed nødt til at forudsætte, at M er konstant gennem hele processen. Dette forbehold har vi taget højde for tidligere, i og med vi har sagt, at processen skal være tidshomogen.

Betydningen af, at afkomstfordelingen skal være poissonfordelt ses bedst ved at tænke på poissonfordelingen som en videre udvikling af binomialfordelingen.

$$P_k = \binom{2N}{k} p^k (1-p)^{(2N-k)} \quad (k = 0, 1, 2, \dots)$$

Binomialfordelingen vil gå mod en poissonfordeling for $2N \rightarrow \infty$ og $p \rightarrow 0$, således at $2N \cdot p$ går mod en konstant.

Hvis vi antager, at afkomstfordelingen skal være binomial-

fordelt, skal det være sådan, at hvert enkelt mutant gen skal have en binomialfordeling som afkomstfordeling. Binomialfordelingen kan forstås som, at hvert mutant gen får $2N$ efterkommere, hvor p_k så giver sandsynligheden for, at der udaf de $2N$ fødte efterkommere er k , som overlever. Poissonfordelingen vil ligeledes give sandsynligheden for, at k mutante gener overlever i næste generation, men i dette tilfælde vil k varierer fra 0 til uendelig. Hvert enkelt mutant skal derfor være i stand til at føde et uendeligt antal efterkommere. Da p samtidig skal være lille, skal populationen være af en sådan beskaffenhed, at der fødes en stor mængde efterkommere, men det er kun en lille del af disse, som overlever og når modningsalderen. Dette vil f.eks. være tilfældet for fisk og andre mindre, æglæggende dyr.

Forgreningsmodellen udtaler sig om sandsynligheden for en enkelt mutants mulighed for overlevelse. Den skelner skarpt mellem $M > 1$ og $M \leq 1$, idet det mutante gen vil have overlevelsessandsynligheden 0 for $M \leq 1$ og en positiv overlevelsessandsynlighed for $M > 1$.

Det første resultat vil kun være gyldigt, såfremt den population det ene mutante gen optræder i til tiden $t = 0$, er meget stor, da sandsynligheden for, at det mutante gen vil blive fikseret, vil gå mod nul, når antallet af ikke mutanter går mod uendelig. Dette vil vi vende tilbage til i afsnit 3.4, hvor vi sammenligner forgreningsmodellen og diffusionsmodellen med de almindelige populationsteoretiske resultater udfra begrebet matematisk generalisation.

Fisher. (7) sætter overlevelsessandsynligheden lig fikseringssandsynligheden for det mutante gen, Dette kan han for $M > 1$ kun gøre, såfremt antallet af ikke-mutanter gener har et endeligt antal, og de har sandsynligheden 1 for udslettelse. Dvs. at middelantal afkom for ikke-mutante gener skal være mindre eller lig med 1, da disse så altid vil have en udslettelsessandsynlighed på 1. Som hovedantagelse ville man nok vælge at sætte den lig med 1, da dette sikrer en nogenlunde stabil populationsstørrelse indtil mutationen indtræder.

Kojima og Kelleher er i deres artikel interesseret i at udregne fikseringssandsynligheden for mutante gener i den menneskelige race og andre større pattedyr. Det vil for disse populationsarter være karakteristisk, at populationsstørrelsen ændres over generationer, og hvert enkelt nyfødt individ vil have en stor sandsynlighed for at nå modningsalderen. Kojima og Kelleher skriver, at det til dette formål vil være velegnet at bruge den negative binomialfordeling som afkomstfordeling.

$$p_k = \binom{r+k-1}{k} p^r q^k$$

hvor p_k er sandsynligheden for, at et mutant har formeret sig til k mutanter i næste generation. r , p og $q = 1-p$ er karakteristiske parametre til fordelingen.

Den sandsynlighedsfrembringende funktion for den negative binomialfordeling vil være givet ved

$$\begin{aligned} G(x) &= \sum_{k=0}^{\infty} p_k x^k = \sum_{k=0}^{\infty} x^k \binom{r+k-1}{k} p^r q^k \\ &= p^r \sum_{k=0}^{\infty} \binom{r+k-1}{k} (xq)^k \end{aligned}$$

For at komme videre benyttes en lille omskrivning

$$\begin{aligned} \sum_{k=0}^{\infty} p_k &= 1 \\ \sum_{k=0}^{\infty} \binom{r+k-1}{k} p^r q^k &= 1 \\ \sum_{k=0}^{\infty} \binom{r+k-1}{k} q^k &= \frac{1}{(1-q)^r} \\ \sum_{k=0}^{\infty} \binom{r+k-1}{k} (qx)^k &= \frac{1}{(1-qx)^r} \end{aligned}$$

og vi får

$$G(x) = p^r \frac{1}{(1-xq)^r}$$

$$G(x) = \left(\frac{p}{1-xq} \right)^r$$

Ved at differentiere $G(x)$ henholdsvis 1 og 2 gange og sætte $x=1$, kan vi få et udtryk for middelværdien og variansen for afkomstfordelingen.

$$\mu = G'(1) = rq/p$$

$$\delta^2 = G''(1) + G'(1) - (G'(1))^2 = rq/p^2 + \mu/p$$

Vi finder igen udslettelsessandsynligheden ved at finde den mindste positive rod til ligningen

$$G(v) = v$$

og den evige overlevelsessandsynlighed u er lig 1 minus udslettelsessandsynligheden. Bartlett fandt i 1955 (2) en approkimerende løsning for v udtrykt ved middelværdien og variansen

$$v = \exp(-2(\mu-1)/\delta^2)$$

og dermed bliver u lig med

$$u = 1 - \exp(-2(\mu-1)/\delta^2)$$

Ved at sammenligne de to forskellige udtryk for den evige overlevelsessandsynlighed fandt Kojima og Kelleher frem til, at for samme middelværdi for de to fordelinger, vil overlevelsessandsynligheden under den negative binomialfordeling altid være mindre end den under poissonfordelingen.

In his comments on Fisher's treatment of the survival probability of mutations, Haldane (1939) pointed out that the use of the Poisson distribution for family size might not be adequate in the case of human populations. The variance of the family size distribution is often in the vicinity of twice the mean of the distribution. For such cases a negative binomial distribution with $p \approx 1/2$ fits actual data fairly well as shown in figure 1. The ultimate survival probability obtained by (7) under the negative binomial assumption is about two-thirds the value computed under the Poisson assumption for genes with small c . It may be, therefore, concluded that the ultimate survival probability given by Fisher (1930) tends to be too large when such populations as man are concerned.

(16)

Den negative binomialfordeling kan fortolkes som, at den giver sandsynligheden (sammenlign binomialfordelingen) for,

at der udaf $(r+k-1)$ fødte er k , der overlever. Dvs., at for forskellige værdier af r vil populationen have forskellig beskaffenhed. r kan derfor bestemmes til at gælde for den menneskelige race. Rent faktisk bestemmes de faste parametre udfra, at man kender middelværdien og variansen for en eller anden populations afkomstfordeling. Fortolkningen af den negative binomialfordeling er ikke helt korrekt, idet den mere præcist udtaler sig om sandsynligheden for, at der er k , som overlever forud for, at der er r døde

Den negative binomialfordeling vil være mere generel end poissonfordelingen, idet man ved at kende middelværdien og variansen kan bestemme de karakteristiske parametre p og r . Endvidere kan middelværdien og variansen være forskellige i den negative binomialfordeling, til forskel for poissonfordelingen, hvor de er lig hinanden

2.4 Forgreningsprocessernes historie. (12), (8)

Vi skal i dette afsnit beskæftige os med, hvorvidt Fisher og Haldane's udledning af fikseringssandsynligheden for mutante gener producerede ny matematisk erkendelse.

Vi har specielt undersøgt om udledningen har medført en udvidelse af teorien for forgreningsprocesser, og afsnittet er derfor hovedsageligt bygget op omkring forgreningsprocessernes historie til og med henholdsvis Haldane's artikel i 1927 og Fisher's i 1930.

Historien omhandler hovedsageligt problemerne med at nå frem til grundsætningen for forgreningsprocesser, hvilken vi derfor kort skal resumere:

Udslettelsen af en population sker med sandsynligheden δ , såfremt middelantal afkom i populationen er mindre end eller lig med 1, mens der altid vil være en positiv sandsynlighed for overlevelse, såfremt middelantal afkom er større end 1.

Selve ordet forgreningsproces (eng. branching process) blev første gang brugt af russerne A. N. Kolmogorov og N. A. Dmitriev i 1947 som betegnelse for de stokastiske processer, der fremkom, når man indførte sandsynlighedsteori til beskrivelse af, hvordan populationsstørrelser vokser eller aftager. De

grundlæggende ideer indenfor forgreningsprocesserne er imidlertid langt ældre end ordet selv.

I 1873 kunne man i et nummer af et engelsk, matematisk tidsskrift kaldet Educational Times læse følgende matematiske problem:

En stor gruppe mennesker - af hvilke vi kun vil beskæftige os med de voksne mænd, som er N i antal, og som alle bærer forskellige efternavne - koloniserer et område. I hver generation er det sådan, at a_0 procent af de voksne mænd ikke får nogen børn af hankøn, a_1 procent af de voksne mænd får et barn af hankøn, a_2 procent får to børn osv. op til 5.

- Find hvor stor en del af efternavnene, der vil være udslukket efter r generationer.

- Hvor mange tilfælde vil der være af samme efternavn, som bæres af m personer.

Problemet var blevet stillet af englænderen Francis Galton udfra hans interesse i, hvorfor mange engelske adelsnavne med tiden var blevet udslettet på trods af, at folketallet i England som helhed var vokset.

Galton modtog kun et svar på sin opgave, og det svar har åbenbart ikke tilfredsstillet ham. I (12) citeres Galton for at have erklæret, at personen havde lavet "a frightful mull of it".

I stedet kontaktede Galton derfor en vis pastor Henry William Watson, som udover sin pastorgerning også excellerede meget i matematik og bjergbestigning. Det er udfra den efterfølgende brevveksling mellem Watson og Galton, at man i dag bedømmer de to til at være de egentlige pionerer indenfor forgreningsprocesser.

Watson angreb Galtons problem ved hjælp af iteration af sandsynlighedsfrembringende funktioner. Han indså, at hvis q_n er sandsynligheden for udslettelse før den n'te generation, så er

$$q_1 = p_0$$

$$q_{n+1} = f(q_n)$$

hvor f er den sandsynlighedsfrembringende funktion og p_0, p_1, p_2, \dots er sandsynligheden for at én voksen mand får henholdsvis 0, 1, 2, ... efterkommere af hankøn. Slutteligt kom Wat-

son frem til ligningen

$$q_\infty = f(q_\infty)$$

som vi også har brugt i afsnit 2.2

Denne ligning vil altid have roden 1 (se afsnit 2.1). På dette grundlag sluttede Watson ukorrekt det uundgåelige i efternavnenes udryddelse.

Længere kom man så ikke i den omgang. Galton prøvede forgæves at få Watson til at fortsætte arbejdet, men han var travlt optaget med andre ting, og man skal så helt frem til 1927 og Haldanes artikel, før nye fremskridt bliver gjort. Vi vil imidlertid lade Haldanes artikel hvile et øjeblik og i stedet springe et par år ekstra frem til 1929.

Danskeren Agner Krarup Erlang publicerede i Matematisk tidsskrift en opgave af tilsvarende karakter, som den Galton havde publiceret i 1873, selvom man ikke mener, at Erlang havde noglet kendskab til Galton og Watsons tidligere arbejde. Erlang havde selv en delvis løsning på sit opstillede problem. I modsætning til Watson indså Erlang således, at ligningen udover roden 1 vil have endnu en rod i intervallet $[0,1]$, såfremt middelantal afkom er større end 1.

Det helt detaljerede og klare bevis for basis-grundsætningen kom dog først, da J. F. Steffensen offentliggjorde sin løsning på Erlangs problem i Matematisk Tidsskrift i 1930. Heri gjorde han det klart, at såfremt middelantal afkom er større end én, er det den mindste af de to rødder, der er den korrekte. Således vil udslettelse finde sted med sandsynligheden 1 for populationer, hvor middelantal afkom er mindre end eller lig med 1, mens der altid vil være en positiv sandsynlighed for overlevelse, såfremt middelantal afkom er større end 1.

Denne grundsætning kan man imidlertid også læse Haldanes artikel fra 1927:

" $x = f(x)$ has two and only two real positive roots, one equal to unity, the other lying between 0 and 1, but near the latter value if k be small (k er selektionskoefficienten). Hence any advantageous dominant factor (svarer til $m > 1$) which has once appeared has a finite chance of survival, how-

ever large the population might be."

Erlangs og Steffensens arbejde foregik helt uafhængigt af Haldane og Fishers arbejde på samme tid, og det på trods af, at problemerne og løsningerne er næsten enslydende.

Det er således helt klart, at Haldane havde kendskab til grundsætningen. At det er Steffensen, der hovedsageligt tages til indtægt for denne, er således kun udtryk for, at han var den første, som offentliggjorde en samlet, komplet analyse af udslettelsessandsynlighederne.

3. DIFFUSIONSMODELLEN.

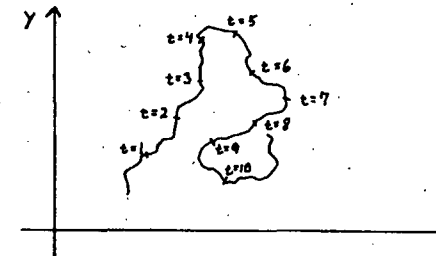
I dette kapitel vil vi i afsnit 3.3 udlede et generelt udtryk for en mutants sandsynlighed for fiksering. Til at opnå dette resultat anvendes Kolmogorovs bagudrettede differential-ligning. I afsnit 3.2 ser vi på, hvorledes Kolmogorovs frem- og bagudrettede differentiaalligninger udledes. Men inden da vil vi kort beskrive dele af diffusionsprocessernes historie.

3.1. Fragmenter af diffusionsprocessernes historie. (10), (20)

I dette afsnit vil vi skitsere udviklingen indenfor diffusionsprocesser. Denne må ses i sammenhæng med rent konkrete problemstillinger, især indenfor fysik. Disse blev man i stand til at løse efterhånden som udviklingen af diffusionsprocesserne skred frem.

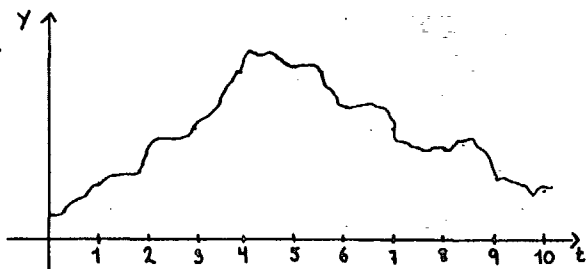
Begrebet diffusion kan typisk forklares ved eksemplet den Brownske bevægelse. I 1827 opdagede den engelske botaniker Robert Brown, at plantepollen, der befinder sig på en væskeoverflade bevæger sig omkring med meget uregelmæssige bevægelser. Brown troede, at han havde opdaget aktive molekyler i organiske og uorganiske legemer.

Fig 3.1.A. Den Brownske bevægelse i to dimensioner



Ved hjælp af Wienerprocessen kan man formelt beskrive den Brownske bevægelse i en eller flere aksers retninger.

Fig 3.1.B. Wienerprocessen (beskrivelse af den Brownske bevægelse i y-aksens retning).



En mere gennemgående teoretisk og kvantitativ beskrivelse af den Brownske bevægelse blev første gang givet af Albert Einstein i 1905. De der for alvor formaliserede beskrivelsen af den Brownske bevægelse var P. Levy og N. Wiener i henholdsvis 1948 og 1923. Den stokastiske proces, som beskriver bevægelsen kaldes også Wienerprocessen.

I løbet af 1930'erne udledte A.N. Kolmogorov (sovjetisk matematiker) to ligninger

$$\frac{\partial}{\partial t} \varphi(x_0, x, t) = \frac{1}{2} \left(\frac{\partial^2}{\partial x^2} (\sigma^2(x) \varphi(x_0, x, t)) - \frac{\partial}{\partial x} (\mu_0(x) \varphi(x_0, x, t)) \right)$$

$$\frac{\partial}{\partial x} \varphi(x_0, x, t) = \frac{1}{2} \sigma_0^2(x) \frac{\partial^2}{\partial x^2} \varphi(x_0, x, t) + \mu_0(x) \frac{\partial}{\partial x_0} \varphi(x_0, x, t)$$

hvor $\sigma^2(x)$ er en infinitesimal varians, $\mu_0(x)$ en infinitesimal middelværdi og $\varphi(x_0, x, t)$ en tæthedsfunktion. De to ligninger kaldes henholdsvis den fremadrettede- og bagudrettede ligning.

Udviklingen af disse ligninger skal ses i sammenhæng med Kolmogorovs interesse for analytiske metoder indenfor sandsynlighedsteori.

De forskellige muligheder for generalisation betyder, at diffusionsprocesser kan anvendes til at beskrive stokastiske fænomener indenfor en række fagområder. I fysikken kan der f.eks. være tale om modeller for, hvordan varme udbreder sig i metal. I biologien kan udover pollenbevægelse i væsker nævnes plante- og dyrespredning. I populationsgenetikken anvendtes diffusionsprocesserne til en sandsynligheds-teoretisk beskrivelse af mu-

tante-geners spredning i en population, for første gang af M. Kimura i 1957. Dette vil vi se nærmere på i det følgende, specielt hvad det betyder for bestemmelse af fikseringssandsynligheden for mutante gener, at man går fra forgreningsprocessens diskrete tid og tilstandsrum til at betragte mutatorers udbredelse som et fænomen med kontinuert tid og tilstandsrum.

3.2. Udledning af Kolmogorovs differentiaalligninger.

Vi vil starte med at se på et specialtilfælde af Kolmogorovs differentiaalligninger, hvorefter vi udleder mere almentgyldige ligninger.

En diffusionsproces er en Markovproces, der har kontinuert tid, tilstandsrum og udfaldsfunktion. En diffusionsproces kan frembringes udfra en diskret Markovproces ved at lade tids-skridtene Δt og de tilhørende tilstandsskridt Δx gå mod 0 på passende måde. I grænsen fremkommer da den kontinuerte diffusionsproces.

Specialtilfældet er en Wienerproces, hvor der er 50% sandsynlighed for, at processen i et infinitesimale tidsrum bevæger sig enten opad eller nedad med konstante infinitesimale tilstandsskridt. Processen vil således fluktuere om begyndelsestilstanden og kaldes derfor en symmetrisk proces. En symmetrisk Random-Walk kan i grænseovergangen føre til en Wienerproces.

I det generelle tilfælde vil processen ikke have en konstant sandsynlighed 1/2 for at bevæge sig henholdsvis opad og nedad. Springene Δx vil heller ikke have en konstant størrelse.

Ved udledningerne af ligningerne går vi altid fra en diskret proces til en kontinuert proces ved at lade Δx og Δt gå mod 0. Her er vi dog nødt til at indføre visse modifikationer for hvorledes Δx og Δt skal gå mod 0, idet det kræves, at variansen skal gå mod en konstant størrelse under grænseovergangen.

3.2.1. Wienerprocessen.

Vi starter med at se på en diskret symmetrisk Random-Walk, som er defineret ved en stokastisk variabel X_t , hvor X_t er summen af de op til t uafhængige stokastiske variable, der hver

antager værdierne Δx og $-\Delta x$ med sandsynligheden $1/2$.
Vi kan umiddelbart se, at

$$E(\Delta x) = (\Delta x/2 + (-\Delta x)/2) = 0$$

Middelværdien af X_t bliver da summen af middelværdierne på hver enkelt tilvækst, vi har altså

$$E(X_t) = (t/\Delta t) E(\Delta x) = 0 \quad \text{hvor } t/\Delta t \in \mathbb{N}$$

Variansen på en tilvækst er

$$V(\Delta x) = E(\Delta x)^2 - (E(\Delta x))^2 = \Delta x^2/2 + (-\Delta x)^2/2 - 0^2 = \Delta x^2$$

hvoraf fås at

$$V(X_t) = (t/\Delta t) (\Delta x)^2$$

hvilket altså gælder for det diskrete tilfælde.

For at $V(X_t)$ skal have en grænseværdi må

$$\Delta x^2/\Delta t \rightarrow \delta^2 \quad \text{for } \Delta x \rightarrow 0 \text{ og } \Delta t \rightarrow 0$$

hvor δ^2 er en konstant og $\delta^2 > 0$. Nu har vi at

$$V(X_t) = \delta^2 t$$

Vi kan allerede nu give et bud på en fordeling af X_t efter grænseovergangen. Da X_t er en sum af $t/\Delta t$ uafhængige stokastiske variable, som hver har middelværdi 0 og varians $(\Delta x)^2$, kan vi nu ved hjælp af den centrale grænseværdisætning sige, at X_t er normalfordelt med middelværdi 0 og varians $t \cdot \delta^2$.

Vi indfører nu en hjælpefunktion f , der indtil videre kan være en vilkårlig funktion. F.eks. ville funktionen defineret ved

$$f(x) = \begin{cases} 0 & ; x \in]-\infty, a[\\ 1 & ; x \in [a, b] \\ 0 & ; x \in]b, \infty[\end{cases}$$

være smart, i det vi da har

$$E(f(X_t) | X_0 = x_0) = P(X_t \in [a, b] | X_0 = x_0)$$

Indtil videre lader vi imidlertid f være en vilkårlig funktion.

For lige at resumere:

Vi går ud fra en diskret symmetrisk Random-Walk Markovproces, hvor X_t er en stokastisk variabel, som er en sum af $t/\Delta t$ uafhængige stokastiske variable, der hver har en middelværdi på 0 og en varians Δx^2 . I grænseovergangen vil processen konvergere mod en Wienerproces under forbehold at $\Delta x^2/\Delta t$ vil gå imod en konstant for $\Delta x \rightarrow 0$ og $\Delta t \rightarrow 0$

Vi vender nu tilbage til det diskrete tilfælde og vil udregne tilvæksten af middelværdier af en vilkårlig funktion af X_t , vi søger altså en formel for udtrykket

$$3.2.1.B. \quad E^{x_0} f(X_{t+\Delta t}) - E^{x_0} f(X_t)$$

idet vi anvender skrivemåden

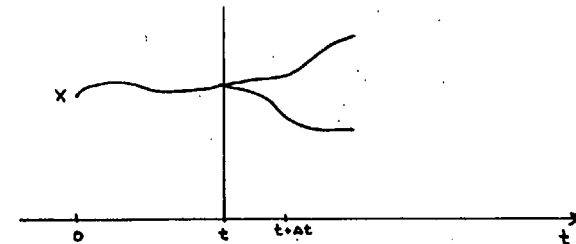
$$E(f(X_t) | X_0 = x_0) = E^{x_0} f(X_t)$$

og er interesserede i at undersøge, hvordan sandsynlighedsfordelingen ændrer sig med tiden. Vi omskriver det første led $E^{x_0} f(X_{t+\Delta t})$ ved hjælp af formlen $E(X) = E(E(X|Y))$. Dette kan gøres på to måder alt efter hvilken del af processens forløb i tiden vi betinger med.

1. mulighed fremadrettet differentiaalligning.

For at udlede den fremadrettede differentiaalligning betinges med forløbet op til tid t .

Fig 3.2.1.C.



$$\begin{aligned}
 E^{x_0} f(X_{t+\Delta t}) &= E^{x_0} (E^{x_0} (f(X_{t+\Delta t}) | \text{forløb op til } t)) \\
 &= E^{x_0} (E(f(X_{t+\Delta t}) | X_t)) \\
 &= E^{x_0} (E^{X_t} (f(X_{t+\Delta t}))) \\
 &= E^{x_0} (\frac{1}{2} f(X_t + \Delta x) + \frac{1}{2} f(X_t - \Delta x))
 \end{aligned}$$

Dette indsættes i 3.2.1.B

$$\begin{aligned}
 E^{x_0} (\frac{1}{2} f(X_t + \Delta x) + \frac{1}{2} f(X_t - \Delta x)) - E^{x_0} f(X_t) \\
 = E^{x_0} (\frac{1}{2} f(X_t + \Delta x) + \frac{1}{2} f(X_t - \Delta x) - f(X_t))
 \end{aligned}$$

ved videre omskrivning benytter vi Taylors formel, og derved forudsætter vi, at f er to gange kontinuert differentiabel og får

$$\begin{aligned}
 E^{x_0} (\frac{1}{2} f(X_t) + \frac{(\Delta x)^2}{4} f''(X_t) + \text{restled}) \\
 + \frac{1}{2} f(X_t) + \frac{(-\Delta x)^2}{4} f''(X_t) + \text{restled} - f(X_t) \\
 = E^{x_0} (\frac{(\Delta x)^2}{2} f''(X_t) + 2 \text{ restled})
 \end{aligned}$$

vi har dermed

$$E^{x_0} f(X_{t+\Delta t}) - E^{x_0} f(X_t) = E^{x_0} (\frac{(\Delta x)^2}{2} f''(X_t) + 2 \text{ restled})$$

$$\frac{E^{x_0} f(X_{t+\Delta t}) - E^{x_0} f(X_t)}{\Delta t} = E^{x_0} (\frac{(\Delta x)^2}{2 \Delta t} f''(X_t) + \frac{2 \text{restled}}{\Delta t})$$

vi er interesserede i grænseovergangen, hvor $\Delta x \rightarrow 0$ og $\Delta t \rightarrow 0$. De to restled vil under den betragtede grænseovergang gå mod nul, og vi har da

$$3.2.1.D \quad \frac{\partial}{\partial t} E^{x_0} f(X_t) = E^{x_0} \frac{\partial^2}{\partial x^2} f(X_t)$$

Vi forudsætter nu, at X_t har tæthedsfunktionen $x \rightarrow \varphi(x_0, x, t)$ givet at $X_0 = x_0$, så er

$$E^{x_0} f(X_t) = \int_{-\infty}^{\infty} f(x) \varphi(x_0, x, t) dx$$

Dette indsættes i 3.2.1.D.

$$\frac{\partial}{\partial t} \int_{-\infty}^{\infty} f(x) \varphi(x_0, x, t) dx = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2} f(x) \varphi(x_0, x, t) dx$$

Vi integrerer højresiden partielt og får

$$\left[\frac{\partial^2}{\partial x^2} f(x) \varphi(x_0, x, t) \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2} f(x) \frac{\partial}{\partial x} \varphi(x_0, x, t) dx$$

Vi vælger f således, at f er nul udenfor et endeligt interval, og højreleddet integreres endnu engang partielt

$$\begin{aligned}
 - \left[\frac{\partial^2}{\partial x^2} f(x) \frac{\partial}{\partial x} \varphi(x_0, x, t) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2} f(x) \frac{\partial^2}{\partial x^2} \varphi(x_0, x, t) dx \\
 = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2} f(x) \frac{\partial^2}{\partial x^2} \varphi(x_0, x, t) dx
 \end{aligned}$$

Alt i alt fås

$$\frac{\partial}{\partial t} \int_{-\infty}^{\infty} f(x) \varphi(x_0, x, t) dx = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2} f(x) \frac{\partial^2}{\partial x^2} \varphi(x_0, x, t) dx$$

Da denne ligning gælder for alle f med kompakt støtte og som er to gange kontinuerte differentiabel, har vi

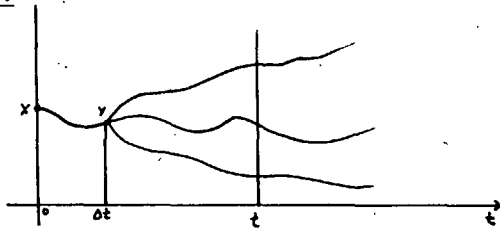
$$\frac{\partial}{\partial t} \varphi(x_0, x, t) = \frac{\partial^2}{\partial x^2} \varphi(x_0, x, t)$$

Dette er den fremadrettede differentialligning for en Wiener-proces, som giver et udtryk for, hvordan sandsynligheds-tætheden ændrer sig med tiden, når man betinger med tilstanden til tid t.

2. mulighed; bagudrettede differentialligning.

For at udlede den bagudrettede differentialligning betinges med forløbet op til Δt .

Fig 3.2.1.E.



$$\begin{aligned}
 & E^{x_0} f(X_{t+\Delta t}) = E^{x_0} (E^{x_0} (f(X_{t+\Delta t}) | \text{forløb op til } \Delta t)) \\
 & = E^{x_0} (E(f(X_{t+\Delta t}) | X_{\Delta t})) \\
 & = E^{x_0} (E^{X_{\Delta t}} (f(X_t))) \\
 & = \frac{1}{2} E^{x_0 + \Delta x} f(X_t) + \frac{1}{2} E^{x_0 - \Delta x} f(X_t)
 \end{aligned}$$

Dette indsættes i 3.2.1.B.

$$\frac{1}{2} E^{x_0 + \Delta x} f(X_t) + \frac{1}{2} E^{x_0 - \Delta x} f(X_t) - E^{x_0} f(X_t)$$

Vi benytter igen Taylors formel og får

$$\begin{aligned}
 & \frac{1}{2} E^{x_0} f(X_t) + \frac{1}{2} \frac{\partial^2}{\partial x^2} E^{x_0} f(X_t) + \frac{1}{4} \frac{\partial^4}{\partial x^4} E^{x_0} f(X_t) + E^{x_0} \text{restled} \\
 & + \frac{1}{2} E^{x_0} f(X_t) + \frac{1}{2} \frac{\partial^2}{\partial x^2} E^{x_0} f(X_t) + \frac{1}{4} \frac{\partial^4}{\partial x^4} E^{x_0} f(X_t) + E^{x_0} \text{restled} - E^{x_0} f(X_t) \\
 & = \frac{(\Delta x)^2}{2} \frac{\partial^2}{\partial x^2} E^{x_0} f(X_t) + 2 \text{restled}
 \end{aligned}$$

Vi har dermed

$$\frac{E^{x_0} f(X_{t+\Delta t}) - E^{x_0} f(X_t)}{\Delta t} = \frac{1}{2} \frac{\Delta x^2}{\Delta t} \frac{\partial^2}{\partial x^2} E^{x_0} f(X_t) + \frac{2 \text{restled}}{\Delta t}$$

Vi ser på grænseovergangen $\Delta t \rightarrow 0$

$$\frac{\partial}{\partial t} E^{x_0} f(X_t) = \frac{1}{2} \delta^2 \frac{\partial^2}{\partial x^2} E^{x_0} f(X_t)$$

Vi forudsætter igen at X_t har tæthedsfunktionen $x \mapsto \varphi(x_0, x, t)$ givet at $X_0 = x_0$

$$\begin{aligned}
 \frac{\partial}{\partial t} \int_{-\infty}^{\infty} f(x) \varphi(x_0, x, t) dx &= \frac{\partial^2}{\partial t^2} \int_{-\infty}^{\infty} f(x) \frac{\partial^2}{\partial x^2} \varphi(x_0, x, t) dx \\
 \frac{\partial}{\partial t} \varphi(x_0, x, t) &= \frac{\delta^2}{2} \frac{\partial^2}{\partial x^2} \varphi(x_0, x, t)
 \end{aligned}$$

Dette var så udledningen af den bagudrettede differentiaalligning for en Wiener-proces, som giver et udtryk for, hvordan sandsynligheds tætheden ændres med tiden, når man betinger med tilstanden til tidst.

3.2.2. Generelle tilvækster.

Som før skal vi undersøge hvordan sandsynligheds-fordelingen ændrer sig med tiden og vi skal igen betragte den fremadrettede og bagudrettede differentiaalligning. Fremgangsmåden vil derfor være den samme.

1. mulighed; fremadrettede differentiaalligning.

Vi har af 3.2.1.B.

$$3.2.2.A. \frac{\partial}{\partial t} (E^{x_0} f(X_{t+\Delta t}) - E^{x_0} f(X_t)) = \frac{\partial}{\partial t} (E^{x_0} E^{X_t} f(X_{\Delta t}) - E^{x_0} f(X_t))$$

Vi sætter nu $X_t = x$ og regner på udtrykket $E^{x_0} f(X_{\Delta t})$ ved brug af Taylors formel.

$$\begin{aligned}
 & E^{x_0} f(x + (X_{\Delta t} - x)) \\
 & = E^{x_0} (f(x) + f'(x)(X_{\Delta t} - x) + \frac{1}{2} f''(x)(X_{\Delta t} - x)^2 + \text{restled}) \\
 & = f(x) + f'(x) E^{x_0} (X_{\Delta t} - x) + \frac{1}{2} f''(x) E^{x_0} (X_{\Delta t} - x)^2 + \text{restled}
 \end{aligned}$$

Vi sætter

$$\begin{aligned}
 E^{x_0} (X_{\Delta t} - x) &= \mu(x) \\
 E^{x_0} (X_{\Delta t} - x)^2 - (E^{x_0} (X_{\Delta t} - x))^2 &= V^{x_0} (X_{\Delta t} - x) = \delta^2(x)
 \end{aligned}$$

og får

$$f(x) + f'(x)\mu(x) + \frac{1}{2} f''(x)(\delta^2(x) + \mu(x)^2) + \text{restled}$$

Vi indsætter nu i 3.2.2.A og har dermed

$$\Delta t [E^{X_0} (f(X_t) + f'(X_t) \mu(X_t) + \frac{1}{2} f''(X_t) (\delta^2(X_t) + \mu(X_t)^2)) - E^{X_0} f(X_t)]$$

Da vi har betragtet en diskret Markov-proces vil vi ved at lade Δt gå mod 0 tilnærme os en kontinert proces.

$$\frac{\delta^2(X_t)}{\Delta t} \longrightarrow \delta_0^2(X_0) \text{ for } \Delta t \rightarrow 0$$

$$\frac{\mu(X_t)}{\Delta t} \longrightarrow \mu_0(X_0) \text{ for } \Delta t \rightarrow 0$$

$$\frac{\mu^2(X_t)}{\Delta t} = \mu(X_0) \frac{\mu(X_t)}{\Delta t} \rightarrow 0 \text{ når } \Delta t \rightarrow 0$$

Hvor $\delta_0^2(X_t)$ er en infinitesimal varians og $\mu_0(X_t)$ er en infinitesimal middelværdi. Disse to værdier indsættes i udtrykket ovenover, og der fås

$$E^{X_0} (\mu_0(X_0) f'(X_0) + \frac{1}{2} (\delta_0^2(X_0) \cdot f''(X_0)))$$

Vi lader X_t have tæthedsfunktionen $x \mapsto \varphi(x_0, x, t)$ givet $X_0 = 0$, og har

$$\int_{-\infty}^{\infty} (\mu_0(x) f'(x) + \frac{1}{2} (\delta_0^2(x)) f''(x)) \varphi(x_0, x, t) dx$$

Dette udtryk opdeles i 1 og 2, således at

$$1 \int_{-\infty}^{\infty} (\mu_0(x) f'(x)) \varphi(x_0, x, t) dx$$

$$= [f(x) (\mu_0(x) \varphi(x_0, x, t))]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x) \frac{\partial}{\partial x} (\mu_0(x) \varphi(x_0, x, t)) dx$$

$$= \int_{-\infty}^{\infty} f(x) (- \frac{\partial}{\partial x} (\mu_0(x) \varphi(x_0, x, t))) dx$$

$$2 \int_{-\infty}^{\infty} \frac{1}{2} (\delta_0^2(x)) f''(x) \varphi(x_0, x, t) dx$$

$$= [f'(x) (\frac{1}{2} \delta_0^2(x) \varphi(x_0, x, t))]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'(x) \frac{\partial}{\partial x} (\frac{1}{2} \delta_0^2(x) \varphi(x_0, x, t)) dx$$

$$= [-f(x) \frac{\partial}{\partial x} (\frac{1}{2} \delta_0^2(x) \varphi(x_0, x, t))]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f(x) \frac{\partial^2}{\partial x^2} (\frac{1}{2} \delta_0^2(x) \varphi(x_0, x, t)) dx$$

$$= \int_{-\infty}^{\infty} f(x) \frac{\partial^2}{\partial x^2} (\frac{1}{2} \delta_0^2(x) \varphi(x_0, x, t)) dx$$

$$1 \cdot 2 \int_{-\infty}^{\infty} f(x) (\frac{\partial^2}{\partial x^2} (\frac{1}{2} \delta_0^2(x) \varphi(x_0, x, t)) - \frac{\partial}{\partial x} (\mu_0(x) \varphi(x_0, x, t))) dx$$

Vi har dermed

$$\int_{-\infty}^{\infty} f(x) \frac{\partial}{\partial t} \varphi(x_0, x, t) dx = \int_{-\infty}^{\infty} f(x) (\frac{\partial^2}{\partial x^2} (\frac{1}{2} \delta_0^2(x) \varphi(x_0, x, t)) - \frac{\partial}{\partial x} (\mu_0(x) \varphi(x_0, x, t))) dx$$

$$\frac{\partial}{\partial t} \varphi(x_0, x, t) = \frac{1}{2} \frac{\partial^2}{\partial x^2} (\delta_0^2(x) \varphi(x_0, x, t)) - \frac{\partial}{\partial x} (\mu_0(x) \varphi(x_0, x, t))$$

Dette er Kolmogorovs fremadrettede differentialligning.

2. mulighed; bagudrettede differentialligning.

Vi har af 3.2.1.B.

$$3.2.2.B. E^{X_0} f(X_{t+\Delta t}) - E^{X_0} f(X_t) = E^{X_0} E^{X_t} f(X_t) - E^{X_0} f(X_t)$$

Dette udtryk omskrives til

$$E^{X_0} g(X_{\Delta t}) - g(x_0)$$

$$\text{hvor } g(y) = E^y f(X_t)$$

$$g(y) = g(x_0 + (y - x_0))$$

$$= g(x_0) + (y - x_0) g'(x_0) + \frac{1}{2} (y - x_0)^2 g''(x_0) + \text{restled}$$

Dette indsættes i 3.2.2.B.

$$E^{X_0} (g(x_0) + (X_{\Delta t} - x_0) g'(x_0) + \frac{1}{2} (X_{\Delta t} - x_0)^2 g''(x_0) + \text{restled}) - g(x_0)$$

$$= g(x_0) + g'(x_0) E^{X_0} (X_{\Delta t} - x_0) + \frac{1}{2} g''(x_0) E^{X_0} (X_{\Delta t} - x_0)^2 + \text{restled} - g(x_0)$$

$$= g'(x_0) E^{X_0} (X_{\Delta t} - x_0) + \frac{1}{2} g''(x_0) (V^{X_0} (X_{\Delta t} - x_0) + (E^{X_0} (X_{\Delta t} - x_0))^2)$$

Vi dividerer nu med Δt og lader $\Delta t \rightarrow 0$

$$\frac{E^{X_0} (X_{\Delta t} - x_0)}{\Delta t} \longrightarrow \mu_0(x_0) \text{ for } \Delta t \rightarrow 0$$

$$\frac{V^{X_0} (X_{\Delta t} - x_0)}{\Delta t} \longrightarrow \delta_0^2(x_0) \text{ for } \Delta t \rightarrow 0$$

$$\frac{(E^{X_0} (X_{\Delta t} - x_0))^2}{\Delta t} \longrightarrow 0 \text{ for } \Delta t \rightarrow 0$$

Vi har alt i alt

$$3.2.2.c. \frac{\partial}{\partial t} E^{x_0} f(X_t) = g'(x_0) \mu_0(x_0) + \frac{1}{2} g''(x_0) \delta_0^2(x_0) \\ = \mu_0(x_0) \frac{\partial}{\partial x_0} E^{x_0} f(X_t) + \frac{1}{2} \delta_0^2(x_0) \frac{\partial^2}{\partial x_0^2} E^{x_0} f(X_t)$$

Vi lader X_t have tæthedsfunktionen $x \mapsto \varphi(x_0, x, t)$ og $E^{x_0} f(X_t)$ kan bestemmes ved

$$E^{x_0} f(X_t) = \int_{-\infty}^{\infty} f(x) \varphi(x_0, x, t) dx \\ \frac{\partial}{\partial t} \int_{-\infty}^{\infty} f(x) \varphi(x_0, x, t) dx \\ = \mu_0(x_0) \int_{-\infty}^{\infty} f(x) \frac{\partial}{\partial x_0} \varphi(x_0, x, t) dx + \frac{1}{2} \delta_0^2(x_0) \int_{-\infty}^{\infty} f(x) \frac{\partial^2}{\partial x_0^2} \varphi(x_0, x, t) dx$$

Heraf fås

$$\frac{\partial}{\partial t} \varphi(x_0, x, t) = \mu_0(x_0) \frac{\partial}{\partial x_0} \varphi(x_0, x, t) + \frac{1}{2} \delta_0^2(x_0) \frac{\partial^2}{\partial x_0^2} \varphi(x_0, x, t)$$

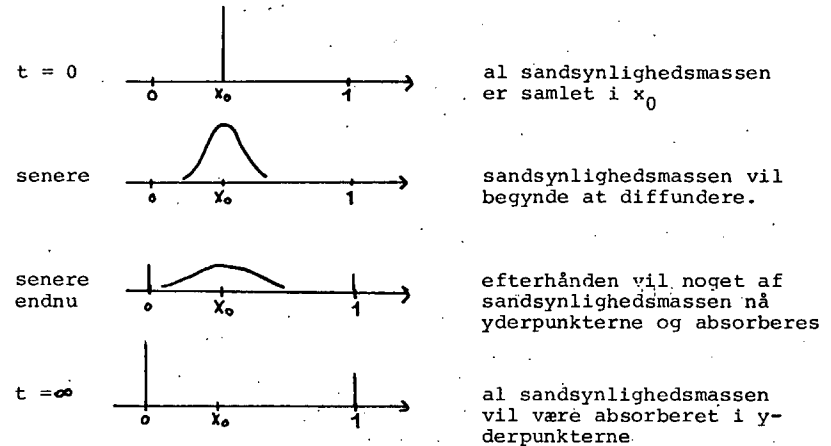
Dette er Kolmogorovs bagudrettede differentiaalligning.

3.3 Gennemgang af diffusionsmodellen.

I forrige afsnit udledte vi den fremad- og bagudrettede differentiaalligning, og vi vil nu anvende den bagudrettede med henblik på at få et udtryk for fikseringssandsynligheden for et mutant gen. Afsnittet er inspireret af Crow (4).

Først vil vi se på tilfældet, hvor genfrekvensen for mutanter ikke systematisk presses i nogen retning, men hvor det kun er tilfældighedsmekanismer, der er årsag til ændringer i genfrekvenserne. I en genetisk terminologi svarer dette til en situation med kun tilfældig genetisk drift og uden selektion. Dernæst ser vi på situationen, hvor genfrekvensen for mutanter presses i en bestemt retning samtidig med, at den også vil være påvirket af tilfældighedsmekanismer.

I det første tilfælde vil processen have en sådan karakter, at den til ethvert tidspunkt vil have en middelværdi svarende til begyndelsesfrekvensen x_0 , og ændringen i middelværdien vil således være nul. Vi kan kort skitsere processens forløb (se også bilag B).



Vi vil nu bruge Kolmogorovs bagudrettede differentiaalligning til at beskrive ændringerne i genfrekvenserne med tiden.

$$\frac{\partial \varphi(x_0, x, t)}{\partial t} = \frac{1}{2} \delta_0^2(x_0) \frac{\partial^2 \varphi(x_0, x, t)}{\partial x_0^2} + \mu_0(x_0) \frac{\partial \varphi(x_0, x, t)}{\partial x_0}$$

Hvor $\delta_0^2(x_0)$ og $\mu_0(x_0)$ er henholdsvis variansen og middelværdien af ændringerne i genfrekvenserne i et infinitesimal tidsinterval.

Når $\mu_0(x_0) = 0$ kan udtrykket forkortes til

$$3.3.A. \frac{\partial \varphi(x_0, x, t)}{\partial t} = \frac{1}{2} \delta_0^2(x) \frac{\partial^2 \varphi(x_0, x, t)}{\partial x_0^2}$$

Problemet bliver så at bestemme $\delta_0^2(x_0)$. Vi kan udtrykke variansen ved x_0 og N_e , hvor N_e er den effektive populationsstørrelse (se bilag C), ved at betragte sandsynlighedsfordelingen for antallet af mutanter til tiden t som en sum af uafhængige stokastiske variable, hvorved fås

$$\delta_0^2(x_0) = x_0(1-x_0)/2N_e$$

En præcis udledning findes i bilag C. Dette indsættes nu i

3.3.A. og der fås

$$\frac{\partial \varphi(x_0, x, t)}{\partial t} = \frac{1}{4N_e} x_0(1-x_0) \frac{\partial^2 \varphi(x_0, x, t)}{\partial x_0^2}$$

Vi vender for et øjeblik tilbage til udtrykket for

$$\frac{\partial}{\partial t} E^{x_0} f(X_t) = \mu_0(x_0) \frac{\partial}{\partial x_0} E^{x_0} f(X_t) + \frac{1}{2} \delta_0^2(x_0) \frac{\partial^2}{\partial x_0^2} E^{x_0} f(X_t)$$

Dette kan omformuleres, ved at se på f i et begrænset interval fra nul til en, og der fås

$$E^{x_0} f(X_t) = \int_0^1 f(x) \varphi(x_0, x, t) + f(0) P^{x_0}(X_t = 0) + f(1) P^{x_0}(X_t = 1)$$

Vi vælger nu f således, at den er tilnærmelsesvis nul i $[0, 1[$ og 1 i punktet 1.

$$3.3.B. \quad E^{x_0} f(X_t) = P^{x_0}(X_t = 1) = u(x_0, t)$$

Vi indsætter nu denne tæthedsfunktion for fikseringssandsynligheden u i Kolmogorovs bagudrettede differentiaalligning, og får

$$3.3.C. \quad \frac{\partial u(x_0, t)}{\partial t} = \frac{1}{2} \delta_0^2(x_0) \frac{\partial^2 u(x_0, t)}{\partial x_0^2} + \mu_0(x_0) \frac{\partial u(x_0, t)}{\partial x_0}$$

Fra 3.3.B. kan vi se, at vi kan finde fikseringssandsynligheden ved at lade $t \rightarrow \infty$

$$u(x_0) = \lim_{t \rightarrow \infty} u(x_0, t)$$

Når $t \rightarrow \infty$ vil $\frac{\partial u(x_0, t)}{\partial t} \rightarrow 0$ og hvis $\mu_0(x_0) \neq 0$ får vi

$$\frac{1}{4\mu_2} x_0(1-x_0) \frac{\partial^2 u(x_0)}{\partial x_0^2} = 0 \quad \Leftrightarrow$$

$$\frac{1}{4\mu_2} x_0(1-x_0) = 0 \quad \vee \quad \frac{\partial^2 u(x_0)}{\partial x_0^2} = 0$$

Det første led bliver aldrig 0 for $x_0 \in]0, 1[$, og derfor er det kun nødvendigt at se på det sidste led

$$\frac{\partial^2 u(x_0)}{\partial x_0^2} = 0$$

$$\frac{\partial u(x_0)}{\partial x_0} = a$$

$$u(x_0) = ax_0 + b$$

Her ved vi, at $u(0) = 0$ og $u(1) = 1$, og ved at indsætte disse værdier i ovenstående udtryk fås

$$u(x_0) = x_0$$

I det tilfælde, hvor processen systematisk presses i en bestemt retning vil $\mu_0(x_0)$ være forskellig for 0. Vi vælger her en ny skala $y = y(x_0)$ således at $\mu_0(y(x_0)) = 0$. For at dette kan lade sig gøre skal y være bijektiv

Gammel skala

$$\begin{array}{c} \xrightarrow{\quad} \\ 0 \quad x_0 \quad 1 \end{array} \quad u(x_0) = \frac{u(x_0) - u(0)}{u(1) - u(0)}$$

Ny skala

$$\begin{array}{c} \xrightarrow{\quad} \\ y(0) \quad y(x_0) \quad y(1) \end{array} \quad u(x_0) = \frac{y(x_0) - y(0)}{y(1) - y(0)}$$

Vi indsætter den ny skala i udtrykket for $\frac{\partial u(x_0)}{\partial x_0}$

$$\frac{\partial u(x_0)}{\partial x_0} = \frac{\partial u(x_0)}{\partial y} \cdot \frac{dy}{dx_0}$$

$$\begin{aligned} \frac{\partial^2 u(x_0)}{\partial x_0^2} &= \frac{\partial}{\partial x_0} \left(\frac{\partial u(x_0)}{\partial y} \right) \frac{dy}{dx_0} + \frac{d^2 y}{dx_0^2} \cdot \frac{\partial u(x_0)}{\partial y} \\ &= \frac{\partial^2 u(x_0)}{\partial y^2} \left(\frac{dy}{dx_0} \right)^2 + \frac{\partial u(x_0)}{\partial y} \cdot \frac{d^2 y}{dx_0^2} \end{aligned}$$

Dette indsættes i 3.3.C.

$$\begin{aligned} \frac{\partial u(x_0)}{\partial t} &= \frac{1}{2} \delta_0^2(x_0) \left(\frac{\partial^2 u(x_0)}{\partial y^2} \left(\frac{dy}{dx_0} \right)^2 + \frac{\partial u(x_0)}{\partial y} \cdot \frac{d^2 y}{dx_0^2} \right) + \mu_0(x_0) \left(\frac{\partial u(x_0)}{\partial y} \cdot \frac{dy}{dx_0} \right) \\ &= \frac{1}{2} \frac{\partial^2 u(x_0)}{\partial y^2} \delta_0^2(x_0) \left(\frac{dy}{dx_0} \right)^2 + \frac{\partial u(x_0)}{\partial y} \left(\frac{1}{2} \delta_0^2(x_0) \frac{d^2 y}{dx_0^2} + \mu_0(x_0) \frac{dy}{dx_0} \right) \end{aligned}$$

Vi vælger nu $y(x_0)$ således at

$$\frac{1}{2} \delta_0^2(x_0) \frac{d^2 y}{dx_0^2} + \mu_0(x_0) \frac{dy}{dx_0} = 0$$

Hvilket betyder

$$\frac{d^2 y}{dx_0^2} = - \frac{2\mu_0(x_0)}{\delta_0^2(x_0)} \cdot \frac{dy}{dx_0}$$

Denne differentiaalligning løses

$$\frac{dy}{dx} = \exp\left(\int_0^x -\frac{2\mu_0(x)}{\delta_0^2(x)} dx\right) = G(x)$$

$$Y(x_0) = \int_0^{x_0} G(x) dx + C$$

Dette udtryk kan så indsættes i

$$\frac{Y(x_0) - Y(0)}{Y(1) - Y(0)}$$

og der fås

$$u(x_0) = \frac{\int_0^{x_0} G(x) dx}{\int_0^1 G(x) dx}$$

hvor

$$G(x) = e^{-\int_0^x \frac{2\mu_0(x)}{\delta_0^2(x)} dx}$$

Vort problem er nu at bestemme $\mu_0(x_0)$ og $\delta_0^2(x_0)$ og en selektionskoefficient s skal nu indopereres i modellen - denne gang på en helt anderledes måde end da vi konstruerede en model for overlevelsessandsynligheden ved hjælp af teorien for forgreningsprocesser.

Lad os først se på det tilfælde, hvor det recessive gen A_2 har en selektionskoefficiens på s imod sig i en population med den effektive størrelse N_e . Idet x er frekvensen af A_1 , er midelværdi og varians af ændringen i genfrekvensen henholdsvis (se bilag C)

$$\mu_0(x_0) = SX(1-X)$$

$$\delta_0^2(x_0) = \frac{x(1-x)}{2N_e}$$

Vi har nu, at $2\mu_0(x_0)/\delta_0^2(x_0) = 4N_e s$ og $G(x) = e^{-\int_0^x 4N_e s dx} = e^{-4N_e s x}$. Nu er

$$u(x_0) = \frac{e^{-4N_e s x_0} - 1}{e^{-4N_e s} - 1}$$

Hermed har vi udledt et generelt udtryk for fikseringssandsynligheden.

Hvis vi sætter $x_0 = \frac{1}{2N}$ i udtrykket for $u(x_0)$ fås

$$u\left(\frac{1}{2N}\right) = \frac{1 - e^{-2\frac{Nk}{N}s}}{1 - e^{-4N_e s}}$$

Hvis $N_e = N$ reduceres dette til

$$\begin{aligned} u\left(\frac{1}{2N}\right) &= \frac{1 - e^{-2s}}{1 - e^{-4Ns}} \\ &= \frac{1 - \left(\frac{(-2s)^0}{0!} + \frac{(-2s)^1}{1!} + \frac{(-2s)^2}{2!} + \dots\right)}{1 - e^{-4Ns}} \\ &= \frac{2s - \frac{1}{2}(2s)^2 + \dots}{1 - e^{-4Ns}} \end{aligned}$$

og er $|s|$ lille haves

$$u\left(\frac{1}{2N}\right) \approx \frac{2s}{1 - e^{-4Ns}}$$

Denne formel gælder endog for negativ s . Er s positiv og lille og N meget stor haves

$$u\left(\frac{1}{2N}\right) \approx 2s$$

Altså det samme resultat som blev opnået med forgreningsmodellen.

Hvis derimod $N_e \neq N$ har vi når $4N_e s \gg 1$

$$\begin{aligned} u\left(\frac{1}{2N}\right) &\approx 1 - e^{-2\frac{Nk}{N}s} \\ &= 1 - 1 + 2\frac{Nk}{N}s - 2\left(\frac{Nk}{N}s\right)^2 + \dots \end{aligned}$$

Er s lille og positiv og $N_e s \gg 1$, kan dette reduceres til

$$u\left(\frac{1}{2N}\right) \approx 2\frac{Nk}{N}s$$

Men stemmer dette resultat overens med, som vi så tidligere, at $u(x_0) = x_0$, når der ingen selektion er, altså når der er tale om en neutral mutation?

$$u\left(\frac{1}{2N}\right) = \frac{1 - e^{-2\frac{Nk}{N}s}}{1 - e^{-4N_e s}}$$

Eksponentialfunktionen rækkeudvikles i både tæller og nævner og der fås

$$u(\frac{1}{2N}) = \frac{2Nes - \frac{1}{2}(2Nes)^2 + \dots}{4Nes - \frac{1}{2}(4Nes)^2 + \dots} \rightarrow \frac{1}{2U} \text{ for } s \rightarrow 0$$

Svaret er altså ja!

Lad nu s og hs være den selektive fordel for henholdsvis den mutante homozygot og heterozygot. Vi har da (se bilag C)

$$M_0(x_0) = sx(1-x)[h + (1-2h)x]$$

$$D_0^2(x_0) = \frac{x(1-x)}{2N_e}$$

Vi har nu, idet C = N_es og D = 2h-1

$$\int \frac{2M_0(x)}{D_0^2(x)} = 4chx - 2cDx^2$$

$$= 2c(D+1)x - 2cDx^2$$

$$= 2cDx + 2cx - 2cDx^2$$

$$= 2cDx(1-x) + 2cx$$

og får dermed

$$G(x) = e^{-2cDx(1-x) - 2cx}$$

Vi har altså

$$u(x_0) = \frac{\int_0^{x_0} e^{-2cDx(1-x) - 2cx} dx}{\int_0^1 e^{-2cDx(1-x) - 2cx} dx}$$

For et komplet recessivt gen h = 0 eller D = -1 har vi f.eks.

$$u(x_0) = \frac{\int_0^{x_0} e^{-2cx^2} dx}{\int_0^1 e^{-2cx^2} dx}$$

$$\frac{k}{2} = 2c - 2Nes$$

$$u(x_0) = \frac{\int_0^{x_0} e^{-\frac{1}{2}kx^2} dx}{\int_0^1 e^{-\frac{1}{2}kx^2} dx} \quad y^2 = kx^2, y = \sqrt{k}x$$

$$= \frac{\frac{1}{\sqrt{k}} \int_0^{\sqrt{k}x_0} e^{-y^2/2} dy}{\frac{1}{\sqrt{k}} \int_0^{\sqrt{k}} e^{-y^2/2} dy} \quad \begin{matrix} \text{1. } \sqrt{k} \text{ meget} \\ \text{Lille} \end{matrix} \approx \frac{x_0 \sqrt{k} e^{-\frac{y^2}{2}}}{\int_0^{\infty} e^{-y^2/2} dy}$$

Vort problem er nu at bestemme $\int_0^{\infty} e^{-\frac{y^2}{2}} dy$
Vi ved, at hvis f(y) er standardnormalfordelt er

$$\int_{-\infty}^{\infty} f(y) dy = 1 \quad \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \frac{1}{2}$$
$$\int_0^{\infty} e^{-y^2/2} dy = \frac{1}{2} \sqrt{2\pi}$$

altså

$$u(x_0) = \frac{x_0 \sqrt{k}}{\frac{1}{2} \sqrt{2\pi}} = \sqrt{\frac{2x_0^2 k}{\pi}}$$

for $x_0 = \frac{1}{2U}$ fås

$$u(\frac{1}{2U}) = \sqrt{\frac{k}{2U^2 \pi}}$$

Vi sætter N_e = N og k = 4Ns og får

$$u(\frac{1}{2U}) = \sqrt{\frac{2s}{\pi U}}$$

Haldane (8) nåede resultatet $\sqrt{\frac{2s}{\pi U}}$ ved at anvende forgreningsprocesser.

Sættes h = 1/2 fås $M_0(x_0) = \frac{1}{2}sx(1-x)$ og dermed er

$$u(x_0) = \frac{\int_0^{x_0} e^{-2Nesx} dx}{\int_0^1 e^{-2Nesx} dx}$$

dette giver

$$u(x_0) \approx \frac{N_e s}{2}$$

under de samme forudsætninger som tidligere, Dette resultat er netop halvt så stort som ved den første beregning af fikseringssandsynligheden.

3.4. Diffusionsmodellens troværdighed.

I dette afsnit vil vi diskutere diffusionsmodellens troværdighed i forhold til forgreningsmodellen og den almindelige populationsgenetiske teori. Vore vurderinger er foretaget ud fra matematikerens synsvinkel og uden særligt kendskab til eventuelle nyere modelverificerende eller modelfalsificerende forskningsresultater fra populationsgenetikken. Først diskuterer vi en række generelle kritikpunkter (såvel positive som negative). Dernæst ser vi nærmere på de mange approximationer, der foretages i modellen, hvorefter vi vurderer diffusionsmodellen i forhold til begrebet matematisk generalisation.

3.4.1 Generel diskussion af diffusionsmodellens troværdighed.

Ved forgreningsmodellen måtte man forudsætte, at M og dermed s forblev konstant under hele processen. For selektionskoefficienten i diffusionsmodellen gælder samme forudsætning. Her må man også hævde, at det forbehold kun kan være opfyldt under meget idealiserede forhold. I de fleste naturlige populationer er det rimeligt at antage, at s flukterer pga. ændringer i miljøet.

Begge modellerne er baseret på, at man beskriver en populations udvikling i tiden vha. en sandsynlighedsfordeling. En ideel beskrivelse heraf - hvis modellen bedst muligt skal minde om virkeligheden - må være en stokastisk proces med kontinuert tid og diskret tilstandsrum, idet en populationsstørrelse ændres med heltallige værdier til vilkårlige tidspunkter. I forgreningsmodellen var tiden diskret - altså må det ikke betyde noget for beskrivelsen af populationens udvikling, at man så at sige kun "tæller" antallet af individer op til bestemte tidspunkter, f.eks. med en generations mellemrum. I diffusionsmodellen er tilstandsrummet kontinuert, men alligevel sker tilstandsændringerne med en konstant, ikke-infinitesimal størrelse. Dette er i modstrid med Kolmogorovs differentiaalligninger, hvor tilstandsændringen sker infinitesimalt, og for at modellen skal være god, må man altså antage, at sådanne unøjagtigheder er uden betydning

Vi argumenterede tidligere for, at med forgreningsmodellen måtte man forudsætte forskellige ting om ikke-mutanter, for at modellen var rimelig, idet det ikke var uden problemer at tale om en egentlig fiksering, når man regnede i absolut antal individer. I diffusionsmodellen regner man derimod i frekvenser og med en fast øvre grænse, således at når x øges bliver $1-x$ mindre, og når x bliver 1, er mutationen fikseret i alle positioner i populationen.

Vi har omtalt, at diffusionsprocesser kan beskrive mange slags fænomener. Selve specificeringen af processen til en model for fikseringssandsynlighed kommer ind med bestemmelsen af middelværdi og varians. Vi bemærker, at disse er bestemt ud fra den verificerede populationsgenetiske teori. Denne teori har imidlertid et gyldighedsområde, f.eks. skal visse ydre faktorer være til stede for, at s kan være konstant, hvilket vi bemærkede tidligere. Holder vi os imidlertid indenfor dette gyldighedsområde, må middelværdi- og variansbestemmelserne være udmærkede.

3.4.2. Analyse af diffusionsmodellens approximationer.

Selve modellen for fikseringssandsynligheden er baseret på Kolmogorovs differentiaalligninger. Disse svarer til en diffusionsproces, som er en kontinuert approximation til en proces med diskret tid og tilstandsrum, når tids- og tilstandsændringen går mod 0.

Dette må dog ikke forveksles med de approximationer, der foretages i selve modellen. Nogle approximationer foretages for at undgå komplekse - måske endda uløselige - ligninger. F.eks. er $\mu_0(x_0) = sx(1-x)$ en approximation, som gælder for lille $|s|$, idet $\mu_0(x_0)$ er givet ved en rækkeudvikling med stigende potenser af s (se bilag C). Det ville få uoverskuelige konsekvenser for beregningen af fikseringssandsynligheden, hvis man medtog bare et led mere.

Andre approximationer foretages kun for at gøre resultaterne mere overskuelige og enkle. Det grundlæggende resultat

opnåes ud fra $\mu_0(x_0) = sx(1-x)$, altså med forudsætningen, at $|s|$ er lille. Dette angives ikke i Crow (1970). Når $N_e = N$ har vi opnået en udmærket approximation til dette udtryk for $|s|$ lille - s skal højst være 0.01 for at approximationen skal være god (se bilag D fig. 3.4.2.A). For $N_e = N$ foretages yderligere en approximation under forudsætningen s positiv og N stor (se bilag D figur 3.4.2.B.). Vi ser, at for $N = 100$ er approximationen ikke særlig god - N skal være endnu større. Når $N_e \neq N$ foretages også et par approximationer. Her gælder stort set analoge betragtninger (se bilag D figur 3.4.2.C. og 3.4.2.D.). Det mest bemærkelsesværdige er her, at N_e skal være et godt stykke over 100, for at approximationerne bliver gode.

Approximationerne er interessante i forhold til forgreningsmodellen jvf. afsnit 3.4.4. Men det er dog tvivlsomt, om de er særlig nøjagtige, for når der er tale om successive approximationer på forskellige niveauer - først selve differentiaalligningen, så i middelværdiberegningen og til sidst i selve udtrykket for fikseringssandsynligheden - altså sammen i forhold til det samme genstandsområde. Under alle omstændigheder må man være opmærksom på indskrænkningerne i modellens anvendelsesområde, hvis approximationerne benyttes.

3.4.3. Matematisk generalisation. (21).

Siden det 16. århundrede har den matematiske generalisation været en vigtig videnskabelig procedure. Princippet anvendes i udviklingen af erkendelse indenfor matematikken selv og dennes anvendelsesområder.

Der kan næppe angives nogen generel generalisationsprocedure for alle de tilfælde, hvor man står overfor en udvidelse af begreber, men visse overordnede betingelser skal være opfyldt.

Et generaliseret begreb skal f.eks. altid kunne reduceres til det oprindelige, når de oprindelige betingelser foreligger. Det generaliserede begreb og det oprindelige må altså ikke være logisk uforbundne eller logisk udelukke hinanden. Det oprindelige begreb skal være et grænsetilfælde af det generaliserede.

Niels Bohr var den første, der helt bevidst anvendte den matematiske generalisationsprocedure i fysikken. Han indførte betegnelsen "korrespondensprincip", som kan opfattes som en generel lovmæssighed for den moderne fysiks udvikling.

De ældre fysiske teorier, hvis gyldighed for et bestemt (begrænset) område af fysiske fænomener er eksperimentelt godtgjort, ophører de ikke med at være gyldige for dette område, selv om nye og mere generelle teorier udformes. Disse gamle teorier bevarer deres gyldighed for det "foregående" fænomenområde som et grænsetilfælde af de nye teorier.

De nye teoriers matematiske apparat, der indeholder karakteriske parametre, som ikke findes i de gamle teorier, går ved passende værdier af disse parametre over i den gamle teoris matematiske apparat.

Et af de mest typiske eksempler er overgangen fra newtonsk mekanik til relativistisk mekanik. Her går alle den relativistiske mekaniks ligninger glidende og asymptotisk over i den newtonske mekaniks ligninger i de grænsetilfælde, hvor hastigheden går mod nul. Et eksempel

$$\text{Klassisk: } F_R = m_0 \cdot a \quad ; \quad \text{moderne: } F_m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \cdot a \left(1 + \frac{v^2}{c^2}\right)$$

Som det ses gælder det, at $F_m \rightarrow F_R$ for $v \rightarrow 0$

Det er klart, at når det drejer sig om naturvidenskab, finder den matematiske generalisation sted i sammenhæng med empiri.

3.4.4. Generalisationsbegrebet og diffusionsmodellen.

Både diffusions- og forgreningsmodellens resultater er generalisationer af den almindelige populationsgenetik resultater, mens diffusionsmodellen yderligere giver en generalisation af forgreningsmodellens resultater. Vi har illustreret disse sammenhænge i følgende figur

Fig. 3.4.4.A.

Figuren illustrerer de resultater, der er fremkommet angående sandsynligheden for fiksering af ét mutant gen ud fra henholdsvis den almindelige populationsgenetik, forgreningsmodellen og diffusionsmodellen.

Alm. pop. gen. teori	forgreningsmod.	diffusionsmod.
$s < 0$: udtaler sig ikke	$U = 0$ gælder for store pop.størrelser	$U = \frac{1 - e^{-2\frac{N_e}{N}s}}{1 - e^{-4N_e s}}$ gælder for vilkår-pop.størrelser
$s = 0$: $U = \frac{1}{2N}$	$U = 0$ overensstemmelse med alm.pop.gen. teori for N meget stor	$U \rightarrow \frac{1}{2N}$ for $s \rightarrow 0$ overensstemmelse med alm.pop.gen. teori for vilkårligt stort N.
$s > 0$: udtaler sig ikke	$U = 2s - \frac{5}{3}s^2 + \frac{7}{9}s^3 \dots$ $U \approx 2s$ for s lille fås	$U = \frac{1 - e^{-2\frac{N_e}{N}s}}{1 - e^{-4N_e s}}$ for s lille fås det mere generelle resultat $U = 2 \frac{N_e}{N} \cdot s$

Ifølge den almindelige populationsgenetiske teori er fikseringssandsynligheden for et neutralt mutant gen lig startfrekvensen $1/2N$. Dette er plausibelt, eftersom der er tale om et neutralt gen, som må have samme sandsynlighed for at få sine efterkommere fikseret i alle $2N$ positioner i populationen som ethvert andet gen.

Fig 3.4.4.A. viser, at både forgrenings- og diffusionsmodellen generaliserer dette resultat, så det nu også gælder for andre s -værdier end netop for $s = 0$. Forgreningsmodellens resultat for $s=0$ er i overensstemmelse med den almindelige popu-

lationsgenetik. Forgreningsmodellen generaliserer imidlertid den almindelige populationsgenetiske teori ved også at give resultater for $s < 0$ og $s > 0$. Det samme gælder for diffusionsmodellen, hvor $u(1/2N)$ har grænseværdien $1/2N$ for $s \rightarrow 0$. Dette kan faktisk udtrykkes endnu mere generelt, eftersom den almindelige populationsgenetiske teori siger, at for neutrale gener opstået med frekvensen x_0 er $u(x_0) = x_0$ efter samme ræsonnement som for $x_0 = 1/2N$. I diffusionsmodellen gælder det tilsvarende generelt at $u(x_0) \rightarrow x_0$ for $s \rightarrow 0$, hvilket f.eks. kan ses af udtrykket $u(x_0) = x_0 + 2N_e s x_0 (1 - x_0) + \dots$ som vi udleder i bilag E.

Nu kan man naturligvis ikke sige, at modellerne bliver mere troværdige, blot fordi de har generaliseret nogle resultater fra den almindelige populationsgenetiske teori. Men der er jo også tale om en harmoni mellem modellerne og den almindelige populationsgenetiske teori i grænsetilfældet. Vi har altså her et lignende forløb, som det vi så med overgangen fra den newtonske mekanik til den relativistiske mekanik, når man ser på modeller for et konkret genstandsområde. Paralleliteten holder dog ikke, når man går over til at se på vore modellers placering i et fagområde, hvilket vil fremgå af kapitel 4. Men trods alt må man sige, at i forhold til modellernes genstandsområde taler det til deres fordel, at de i grænsetilfældet stemmer overens med den alment accepterede populationsgenetiske teori.

Samtidig må man sige, at diffusionsmodellen på mange områder også fremstår som en generalisation i forhold til forgreningsmodellen. F.eks. ser man af fig. 3.4.4.A., at approximationen $u(1/2N) \approx 2s$ fra forgreningsmodellen generaliseres til $u(1/2N) \approx 2N_e s/N$ i diffusionsmodellen. Den nye models resultat er således en matematisk generalisation, der indeholder den gamle model resultat som et specialtilfælde. Når $N_e = N$ fås jo netop $u = 2s$. Desuden er det ofte sådan, at jo større en population er, jo nærmere vil N_e ligge N . Så vi ser en glidende overgang fra den nye til den gamle model efterhånden som populationens størrelse vokser, og vi får den gamle models resultater i det grænsetilfælde, hvor dens betingelser er opfyldt. Da N_e afviger fra N i langt de fleste naturlige såvel som kontrollerede

populationer, har den nye models resultat langt mere generel betydning. Samtidig kan forholdet N_e/N bestemmes eksperimentielt. For mennesket er N_e/N omkring $2/3$, således at $u(s) = (4/3)s$, hvilket svarer til et resultat fra forgreningsmodellen, når afkomstfordelingen er den negative binomialfordeling (16).

4. MODELLEARNES ANVENDELSE

Vi vil i dette kapitel forsøge at gøre rede for, hvilken betydning de gennemgåede modellers resultater har haft indenfor forskellige fagområder.

Det viser sig, at modellernes resultater finder anvendelse i en i begyndelsen af halvfjerdserne opstået debat omkring den bestående evolutionsteori. Den ene part i debatten benævnes som regel selektionister og den anden neutralister. Blandt andet udfra diffusionsmodellen, men også mange andre udregninger og ræsonnementer, mener neutralisterne at være nået frem til en bedre teori for, hvorledes evolutionen gebårder sig på det molekylære niveau.

Vi vil i afsnit 4.1 give en kort redegørelse for, hvad debatten hovedsageligt drejer sig om, og derefter ligeledes kort gøre rede for, hvordan diffusionsmodellens resultater bruges i debatten.

Yderligere finder modellens resultater anvendelse indenfor plante- og dyreavl, hvor det har løst et mangeårigt problem angående optimering af den genetiske gevinst ved kunstig selektion, hvilket vi ser på i afsnit 4.2.

4.1. Modellernes anvendelse i en debat om evolutionsteorien

Vi vil først redegøre for selektionisternes opfattelse, for dernæst at påpege hvor neutralisterne har en divergerende mening.

Darwins evolutionsteorier var overvejende intuitive forestillinger om, hvordan udviklingen fra art til art kunne finde sted. Siden de Mendelske love så at sige blev genopdaget i begyndelsen af 1900-tallet, og genetikken blomstrede op som fagområde, har darwinisterne gennem observationer, genetiske eksperimenter, logiske ræsonnementer og matematiske modeller arbejdet på, at få den klassiske Darwinisme tilpasset og udviklet i forhold til gen-læren.

Groft sagt har man herved fået den klassiske Darwinisme formuleret i genetik-vendinger: Sker en fordelagtig mutation, dvs. en, der gør individet mere overlevelsesdygtigt, vil den

gennem generationer spredes til hele populationen og erstatter den hidtil dominerende type. Ufordelagtige mutationer elimineres derimod forholdsvis hurtigt. Set over mange generationer er evolutionen således en jævn fremadskridende proces, hvor det hele tiden er det til omgivelserne bedst tilpassede individ, der overlever. (15)

Efter selektionisternes opfattelse afgøres en mutations skæbne således hovedsageligt af, om mutationen er selektiv fordelagtig eller ej. I modsætning hertil står neutralisterne, som mener, at de fleste genfikseringer i naturen finder sted på grundlag af mutationer, der er så tæt på at være selektivt neutrale, at deres skæbne hovedsageligt afgøres af tilfældig genetisk drift. (5)

Et af de væsentligste spørgsmål i debatten er derfor:

Hvilken del af den observerede evolution på det molekylære niveau skyldes naturlig selektion og hvilken del skyldes tilfældig genetisk drift?

I dette spørgsmål har diffusionsmodellen en plads som argument fra neutralisternes side. Udfra denne mener neutralisterne nemlig, at man kan vise, at selektion alene ikke kan være årsag til alle de genfikseringer, der finder sted i naturen, og dermed støttes neutralisternes forestillinger om, at næsten neutrale mutationer spiller en afgørende rolle i evolutionen.

Lad os resummere diffusionsmodellens resultat m.h.t. sandsynligheden for fiksering af én mutation.

$$u\left(\frac{1}{2N}\right) = \frac{1 - \exp(-2N_e s/N)}{1 - \exp(-4N_e s)}$$

som vi tidligere har set kan reduceres til

$$u\left(\frac{1}{2N}\right) \approx 2 \frac{N_e}{N} s$$

såfremt $4N_e s \gg 1$ og $s \ll 1$, dvs., såfremt mutationen har en selektiv fordel, og der er tale om en meget stor effektiv populationsstørrelse.

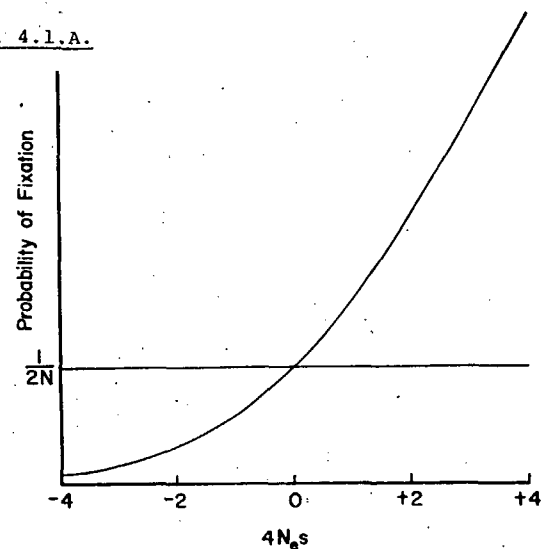
Er $|4N_e s| \ll 1$ har vi i grænsen for $s \rightarrow 0$, at

$$u\left(\frac{1}{2N}\right) = 1/2N$$

På figur 4.1.A. ses kurven for u som funktion af den effek-

tive populationsstørrelse N_e og selektionskoefficienten s (produktet af de to).

Fig. 4.1.A.



Frå kilde (15)

Udfra kurven ses, at neutrale mutationer og endda ufordelagtige mutationer har en positiv sandsynlighed for fiksering, medmindre der er tale om meget ufordelagtige mutationer. Forekommer disse næsten neutrale mutationer således i stort antal i forhold til antallet af fordelagtige mutationer, må man gå ud fra, at de næsten neutrale mutationer spiller en ikke ubetydelige rolle i evolutionen.

Man ser ligeledes, at selvom en mutation er selektiv fordelagtig er det ikke sikkert, den bliver fikseret, medmindre fordelagtigheden er meget stor. Ifølge Kimura (13, s.11) er dette endnu ikke fuldt ud anerkendt af selektionisterne i mange af de evolutionsteoretiske diskussioner.

Rent umiddelbart giver diffusionsmodellen altså god støtte til de ideer, som neutralisterne har omkring evolutionen på det molekylære niveau. Vi vil nu gennemgå to eksempler på, hvordan neutralisterne argumenterer på grundlag af diffusionsmodellen. Det første eksempel skal vise, at selektion ikke

alene kan være årsag til alle de gensubstitutioner, der sker i naturen, og at neutrale gener derfor også må have en betydelig rolle. Det andet eksempel bruges til at forklare den genetiske variation i naturen.

Vi lader K betegne middeltal gensebstitutioner per generation. K er altså et udtryk for, hvor hurtigt evolutionen skrider frem.

Antallet af fordelagtige mutante gener, som optræder i hele populationen, kaldes r_m . Bemærk, at vi kun regner K ud for de selektivt fordelagtige mutationer, det er pointen ved udregningen.

Vi har da

$$4.1.B \quad K = r_m \times u \left(\frac{1}{2N} \right) = 2s(N_e/N) \times r_m$$

i-det vi indsætter approximationen for selektivt fordelagtige gener for $u(1/2N)$.

Vi lader r være mutationsraten per gamet per generation, og har således at $r = r_m/2N$, som vi indsætter i 4.1.B. og får

$$K = 4 \cdot N_e \cdot s \cdot r \quad (13,14)$$

såfremt det kun er fordelagtige mutationer, der fikseres. Vi skal altså holde $4N_e s$ meget større end 1.

Lader vi $N_e = 10^4$ og $s = 0,01$ får vi $K = 400 \cdot r$

Værdien af r kendes ikke, men skal vi støtte os til selektionisternes opfattelse af, at selektivt fordelagtige mutationer findes i et minimalt antal i forhold til antallet af lethale mutationer, kan vi sætte nogle nedre grænser for, hvor stor r kan være. Eksempelvis ved man, at den lethale mutationsrate per kønscelle per generation er 0,015 i bananfluer, og antager man, at fordelagtige mutationer optræder en tusindedel så ofte som lethale, har vi, at $r = 1,5 \cdot 10^{-5}$ mutationer per gamet per generation. Indsat i 4.1.B. har vi

$$K = 400 r = 400 \cdot 1,5 \cdot 10^{-5} = 6 \cdot 10^{-3} \text{ genfikseringer/generation}$$

svarende til én genfiksering for hver 170 generationer.

Selvom udregningen bærer præg af nogle løse antagelser og cirka-værdier, står resultatet i skærende kontrast til det antal gensubstitutioner, man har observeret for forskellige

populationer i naturen. For mennesker har man f.eks. målt en gensubstitutionsrate på 50 genfikseringer per generation, altså 8500 gange så mange fikseringer, som ovenstående resultat giver anledning til. Neutralisternes svar på uoverensstemmelserne mellem de beregnede og observerede gensubstitutionsrater er, at det resultat man får ved kun at regne gensubstitutionsraten ud for fordelagtige mutationer, skal suppleres op med det antal gensubstitutioner, der sker på grundlag af neutrale mutationer; et antal, der ifølge udregningerne, er betragteligt stort.

Som en anden beregning, der har betydning i debatten, og som diffusionsmodellen har givet anledning til, vil vi nævne middeltal generationer det tager for en mutation at blive fikseret, såfremt den bliver det, og middeltal generationer det tager for en mutation at dø ud, såfremt den gør det.

$t_0(x_0)$ er antal generationer indtil tab af det mutante gen, når der ses bort fra eventuelt fiksering.

$t_1(x_0)$ er antal generationer indtil fiksering af det mutante gen, når der ses bort fra eventuelt tab.

Begge er stokastiske variable, hvis middelværdi kan vises at være

$$E(t_0(x_0)) = -\frac{1}{x_0} \{4N_e(1-x_0) \ln(1-x_0)\}$$

$$E(t_1(x_0)) = -\frac{4N_e x_0}{1-x_0} \ln x_0 \quad (4)$$

$$E(t_0(\frac{1}{2N})) = -\frac{1}{\frac{1}{2N}} \{4N_e(1-\frac{1}{2N}) \ln(1-\frac{1}{2N})\}$$

$$= 4N_e(1-2N) \ln(1-\frac{1}{2N})$$

$$\approx 4N_e$$

$$E(t_1(\frac{1}{2N})) = -\frac{4N_e \cdot \frac{1}{2N}}{1-\frac{1}{2N}} \cdot \ln(\frac{1}{2N})$$

$$= \frac{4N_e \cdot \ln(2N)}{2N-1}$$

$$\approx \frac{2N_e}{N} \ln(2N)$$

Da man i slutningen af tresserne med indførslen af elektroforesestudier blev i stand til mere detaljeret at bestemme gensammensætningen i forskellige individer, opstod der visse uoverensstemmelser mellem den teori selektionisterne var kommet frem til og observationerne. Specielt viste den genetiske variation sig at være meget større end, hvad man kunne forvente.

Ifølge neutralisterne forklarer ovenstående resultat den genetiske variation i naturen som værende overgangsstadier for konkurrerende alleller. Eftersom der går temmelig lang tid ($4N_e$ regnet i generationer) inden en allel/mutant bliver fikseret i store populationer, vil 2 eller flere alleller længe kunne opholde sig samtidigt i den samme population og dermed bidrage til den genetiske variation.

Selektionisterne har forklaret den genetiske variation på en anden måde. På grund af vores ringe biologiske indsigt undlader vi at gøre rede for den her, men skal blot konstatere, at forklaringen af den genetiske variation i naturen indtager en vigtig plads i debatten, og at diffusionsmodellen også her giver sit besyv med som argument fra neutralisternes side.

Vi har i afsnittet undladt at uddybe selektionisternes argumenter mod diffusionsmodellen, eftersom de er meget indre biologiske, og ikke kører på det overordnede matematiske plan, som vores modelanalyser har gjort.

Vi har imidlertid set, hvordan neutralisterne bruger diffusionsmodellens resultater i debatten. Det er dog klart, at neutralisterne ikke udelukkende bygger deres argumentation på denne ene model, men at denne skal sidestilles med laboratorieforsøg, observationer i felten m.m.

4.2. Diffusionsmodellens anvendelse i plante- og dyreavl.

I mange år drejede teorien om den kunstige selektion sig om at maksimere den genetiske gevinst per generation.

Problemet med selektionsgrænsen - dvs. den maksimale genetiske forbedring, det var muligt at opnå under en given fremgangsmåde med kunstig selektion - blev udeladt af betragtningerne.

Det var Robertson (1960), som pionerede indenfor udviklingen af en teori for selektionsgrænser baseret på teorien for genfiksering i en endelig population. (15)

Ved en omskrivning af diffusionsmodellens formel for genfikseringssandsynligheder udledte han formlen

$$u(x_0) - x_0 = 2N_e s x_0 (1-x_0) + \dots \quad (\text{se bilag E})$$

hvor $u(x_0) - x_0$ er den maksimale gevinst, der kan opnås ved succesiv (kunstig) selektion; at dette er tilfældet vil vi ikke argumentere nærmere for.

Af formlen indser man imidlertid, at da $s x_0 (1-x_0)$ er middelændringen i genfrekvensen (se bilag C), er den maksimalt opnåelige gevinst netop $2N_e$ gange gevinsten i første generation.

Diffusionsmodellen har således også fundet anvendelse indenfor teorien for plante- og dyreavl.

5. DISKUSSION:

Vi vil i dette kapitel diskutere og undersøge hvilke svar gennemgangen af forgrenings- og diffusionsmodellen giver på spørgsmålene i problemformuleringen. Herefter vil vi komme med nogle generelle betragtninger omkring resultatet af disse undersøgelser, bl.a. vurdere dem i forhold til den almindelige modeldiskussion på IMFUFA.

5.1. Modellernes konstruktion.

Vi har i det forløbne set to forskellige stokastiske processer anvendt på det samme problem. Ved anvendelse af forgreningsprocessen, der arbejder med diskret tid og diskret tilstandsrum, påkøber man en bestemt afkomstfordeling for de mutante gener, for derefter at approximere sig frem til et formelt set mere overskueligt resultat. I diffusionsprocessen, der arbejder med kontinuert tid og kontinuert tilstandsrum, indsættes bestemte udtryk for middelværdien og variansen i Kolmogorovs bagudrettede differentiaalligning, som derefter bliver løst for bestemte randbetingelser. Ligesom i forgreningsmodellen kan løsningen approksimeres til nogle mere overskuelige resultater under visse forudsætninger.

Begge modeller består således af, hvad man kunne kalde en proces-del, som lægger sig tæt op ad den generelle teori for stokastiske processer, samt en parameterbestemmelse, som knytter sig mere eller mindre til den ikke-matematiske teori for genstandsområdet (i forgreningsmodellen drejer det sig om at bestemme en afkomstfordeling, i diffusionsmodellen en middelværdi og en varians). Man approximerer eventuelt til slut for at opnå et mere overskueligt resultat.

Proces-delen består mere præcist af et udtryk for, hvordan en bestemt proces forløber med tiden. I forgreningsprocessen givet ved iteration af en sandsynlighedsfrembringende funktion i diffusionsprocessen som en differentiaalligning. Disse procesdele har ikke nogen specifik anvendelse, men kan benyttes indenfor vidt forskellige fagområder.

Den matematiske teori for processerne ligger klar på forhånd. Modelkonstruktøren skal vælge en til genstandsområdet passende proces og dernæst skal processens parametre formuleres i gen-

standsområdets sprog.

Med hensyn til modellernes konstruktion bemærker man også, at der på intet tidspunkt i modelbygningen forekommer værdier, der skal estimeres. Det nærmeste, man kommer, er i diffusionsprocessen, hvor man indsætter bestemte udtryk for middelværdi og varians. Man skal imidlertid lægge mærke til, at det er et bestemt udtryk, man indsætter, og ikke en bestemt værdi. Dette skyldes naturligvis, at det er et kvalitativt resultat, man er interesseret i, nemlig at finde u som funktion af s .

5.2. Modellernes troværdighed.

Vi har i vores projektarbejde brugt en del af tiden på at klargøre hvilke antagelser, modellerne bygger på, og hvad det er i modelbygningen, der betinger de forskellige antagelser. Vi ved, at der gøres antagelser tre steder under konstruktionsforløbet, nemlig ved procesvalget, ved parameterbestemmelsen samt ved de matematiske approximationer. Mere specifikt:

1) Procesdelen:

Når man vælger den proces, man vil benytte til sit problem, overfører man implicit de antagelser, processen bygger på, til selve modellen. Ved modelkonstruktionen må man derfor sørge for, at de antagelser, processen bygger på, er rimelige for ens genstandsområde. For eksempel bygger forgreningsprocessen på diskret tid. Vælger man forgreningsprocessen til at beskrive udviklingen i antallet af mutante gener, har man hermed antaget, at udviklingen foregår i "hop" af en bestemt tids varighed. Det er klart, at en populations udvikling aldrig vil foregå i konstante spring. Man bliver derfor nødt til at godtgøre/antage, at det er uden betydning for beskrivelsen at regne med diskret tid.

2) Parameterbestemmelsen:

For forgreningsmodellens vedkommende har vi med gennemgangen af (16), der i stedet for poissonfordelingen anvender den negative binomialfordeling, set, at valget af afkomstfordeling har stor betydning for, hvilket resultat man når frem til. Der findes ingen verificeret teori for, hvordan afkomstfordelingen for mutante gener skal se ud, og man har derfor ingen garanti for, at den fordeling man har valgt, er den bedst valgte. Dette er dog ikke ensbetydende med, at man vælger en tilfældig fordeling som

afkomstfordeling, idet man gør sig nogle overvejelser om, hvilke populationer der passer til hvilke fordelinger. For eksempel var Fisher og Haldane udmærket klar over, at poissonfordelingen ikke passer særlig godt til den menneskelige race, hvilket Fisher også gør opmærksom på i (7), hvori han skriver, hvilke betingelser, han mener, populationen skal opfylde for at følge poissonfordelingen som afkomstfordeling.

Den negative binomialfordeling er mere generel end poissonfordelingen, idet man ved at kende middelværdien og variansen for en populations afkomstfordeling kan bestemme parameteren r . Således kan den negative binomialfordeling tilpasses de fleste populationer.

I diffusionsmodellen er situationen lidt anderledes, eftersom det her ikke er nødvendigt at vælge en bestemt afkomstfordeling. Her skal fastsættes et udtryk for middelværdien og variansen i ændringer i genfrekvensen, og det har man teorier for.

Der er altså en forskel på forgrenings- og diffusionsmodellens parameterbestemmelser. I diffusionsmodellen bygger man sit valg på teoribaserede udregninger af middelværdi og varians, mens afkomstfordelingen i forgreningsmodellen er valgt ud fra intuitive forestillinger om, hvordan forskellige populationer formerer sig.

3) Approximationerne:

Mens valget af proces og parameterbestemmelsen betinger nogle rent biologiske antagelser om de populationer, modellerne udtaler sig om, indskrænker de matematiske approximationer modellernes gyldighedsområde m.h.t. populationsstørrelser og størrelsen af s . Det må være et krav til approximationerne, at de er rimelige i det område, hvor man bruger modellens resultater. Vi har set, at især diffusionsmodellen gør meget brug af approximationer på forskellige niveauer, hvilket ikke var nær så udpræget ved forgreningsmodellen.

4) Yderligere kommentarer til modellernes troværdighed:

Vi har endvidere set på modellerne i forhold til begrebet matematisk generalisation. Vi fandt her frem til, at de to modeller rent matematisk korresponderer med hinanden og med den almindelige populationsgenetikens resultater. Hvis man ud fra modellerne havde deduceret andet, end at overlevelsessandsynligheden er lig startfrekvensen, når der ingen selektion er, havde dette

været en modstrid med den alment accepterede populationsgenetik. En sådan komplikation er imidlertid ikke opstået, og modellerne er harmoniske indbyrdes og i forhold til den verificerede teori.

5.3. Modellernes placering i et fagområde.

De modeller, vi har betragtet, producerer nogle kvantitative udtryk, som beskriver nogle forhold i en afgrænset del af virkeligheden - genstandsområdet. Udtrykkene fortolkes og indgår i den evolutionsteoretiske debat - fagområdet.

Modellerne anvendes, sammen med nogle konkrete forsøgsresultater, til at argumentere for bestemte synspunkter, angående hvordan genstandsområdet er sammensat, idet genstandsområdets struktur er afgørende for hele opfattelsen af, hvorledes evolutionen foregår.

Modellerne har altså en vis betydning for, hvorledes teorien for evolutionen udformes. Modellernes rolle i fagområdet er derfor, hvad man kunne kalde teoridannende, idet deres resultater benyttes som et bindeled mellem forsøgsresultater og selve teortudformningen.

5.4. Problemer med modellernes formalside.

Vi har set, at Haldane tilsyneladende var den første, der fandt frem til grundsætningen for forgreningsprocesser.

Watson nåede frem til, at udslettelsessandsynligheden i en forgreningsproces altid er 1. Haldane fandt frem til, at dette var ukorrekt, og nåede derigennem frem til, at mutante gener faktisk har positiv overlevelsessandsynlighed, når middelantal afkom er større end 1. Dette er også et intuitivt mere fornuftigt resultat end konsekvenserne af Watsons forslag, hvis det blev anvendt på samme populationsgenetiske problem.

Haldane måtte altså færdiggøre en del af den matematiske teori for forgreningsprocesser for at kunne konstruere en fornuftig model for genstandsområdet.

Det er bemærkelsesværdigt, at Haldane nærmest kun nævner dette resultat i en bisætning uden nogensomhelst forklaring på, hvordan han nåede frem til det. Måske fandt han det uvigtigt i forhold til sit primære formål: At forklare hvordan evolutio-

nen foregår.

Vi har ikke i vort arbejde med diffusionsmodellen fundet eksempler på, at man har måttet klare hidtil uløste matematiske problemer for at kunne nå videre i modelkonstruktionen.

5.5. Debatten om matematiske modeller.

Der er ingen tvivl om, at de modeller, vi har betragtet er teoriladede. Nu kan man imidlertid tale om forskellige grader af teoriladethed. En del af debatten drejer sig netop om at klassificere matematiske modeller i forhold til dette begreb.

Man må konstatere, at diffusionsmodellen er mere teoriladet end forgreningsmodellen. I diffusionsmodellen bygger bestemmelserne af middelværdi og varians på teori, mens dette ikke er tilfældet, når man skal vælge afkomstfordeling i forgreningsmodellen. I begge modellerne er det dog nødvendigt at kende hovedegenskaberne ved gener, samt at kende begreberne selektion og mutation. I forgreningsmodellen ser man også, at begrebet selektion indføres på en mere intuitiv måde, end tilfældet var i diffusionsmodellen, hvor teorien netop sagde, at s skulle indarbejdes i middelværdien på en bestemt måde.

Visse ad hoc-modeller har nærmest været bestillingsarbejde, der skulle understøtte bestemte politiske og økonomiske interesser. Modellerne vi har set på, udnyttes til at argumentere for bestemte synspunkter i en grundvidenskabelig debat. Det er dog vigtigt at bemærke, at debatten først er opstået efter modellerne er blevet konstrueret. Dette skyldes ikke mindst, at modellerne har det, vi kalder en teoridannende karakter. Troværdigheden af argumenter, der bl.a. bygger på disse modeller, må derfor også vurderes ud fra en diskussion af modellernes troværdighed.

En del af debatten om modeller har drejet sig om, hvorvidt modeller af og til konstrueres for matematikkens/matematikerens skyld, eller om det altid er intentioner om at producere ny grundvidenskabelig erkendelse, der har medført, at modeller er blevet konstrueret. En diskussion heraf ses f.eks. i (18).

Forgrenings- og diffusionsmodellen ser ud til at have haft det sidstnævnte formål. Matematikken har haft en redskabsfunktion og har været af sekundær betydning i forhold til ønsket om at opnå et dybere kendskab til genstandsområdets struktur. Et problem, vi ikke har beskæftiget os med, er hvorvidt anvendelsen af matematik har haft en fremmedgørende effekt overfor populationsgenetikere med ringe kendskab til matematik, og om det-

Et andet aspekt, vi ikke har beskæftiget os med, er den nyere udvikling indenfor modeller for fikseringssandsynligheder. Det er givet, at en del af den kritik, vi har givet af de to modeller, er påpeget enten af modelkonstruktørerne selv eller fra anden side. Dette, i sammenhæng med de stadigt mere raffinerede metoder indenfor sandsynlighedsteorien, har resulteret i forskellige forsøg på forbedringer af modellerne, hvilket vi ikke har fået tid til at gå nærmere ind på. De modelteoretiske diskussioner, vi har haft, har dog ikke været afgørende afhængige af kendskab til den nyeste udvikling indenfor området.

6. KONKLUSION.

Vi har nu analyseret spørgsmålene i problemformuleringen med udgangspunkt i to konkrete modeller.

Det har specielt vist sig at, at det ikke er uden problemer at skulle vurdere troværdigheden af modeller indenfor et naturvidenskabeligt område som matematiker.

Troværdighedsspørgsmålet er hovedsageligt blevet drejet over i retning af, hvilken virkelighed modellerne gælder for, eller hvad der er modellernes gyldighedsområde. Vi har f.eks. konstateret nogle indskrænkninger i modellernes gyldighedsområde p.gr.a. en række betingelser som opstilles såvel i procesvalget som i parameterbestemmelsen.

Kendskabet til modellernes gyldighedsområde er nok så vigtigt for, hvor meget man kan tillade sig at slutte ud fra modellens resultater. Modellernes troværdighed afgøres først og fremmest i sammenhæng med empiri, men dog har vi fundet ud af, at approximationerne er rimelige hver for sig, og at modellerne udgør et konsistent system indbyrdes og i forhold til den verificerede populationsgenetik.

Det har været vigtigt for overskueligheden af analysearbejdet, at forsøge en begrebsmæssig opdeling af modellerne i forskellige dele, således at man opnår en klar fornemmelse af, hvad der er model, og hvad der er virkelighed, og hvilke træk ved modellen, der er formale begreber, og hvilke der viser hen til den virkelighed, der modelleres. Vor opfattelse af begrebet teori-ladethed har haft en vigtig rolle i denne skelnen mellem forskellige bestanddele i modellerne.

Desuden har vi undersøgt modellernes formål, hvor vi bl.a. har set, at modellerne har en interessant placering i spillet mellem forskellige opfattelser af elementer ved evolutionen. Specielt er det værd at bemærke, at modellerne har betydning for teoridannelsen uden egentlig at indgå som en del af teorien.

Alt i alt håber vi, at resultatet af disse undersøgelser kan være med til at videreudvikle kritikken af teoriladede modeller.

APPENDIX

FORMIDLING AF STOKASTISKE MODELLER I POPULATIONSGENETIK TIL GYMNASIET OG HF

Vi har nu set på to stokastiske modeller, der besvarer spørgsmålet, "hvad er sandsynligheden for fiksering af mutante gener?"

Vi har diskuteret, hvordan disse modeller er opbygget. Vi har forsøgt at vurdere, hvorvidt den populationsgenetiske erkendelse, som modellerne producerer, er troværdig - dog ud fra det forbehold, at vi ikke har detaljeret biologisk indsigt. Vi har undersøgt om modelkonstruktionen kan medføre udvikling af ny matematisk erkendelse. Og vi har set på, hvordan modellerne og deres resultater indgår i et fagområde.

Spørgsmålet i dette appendix er nu, om det er relevant og muligt at formidle stokastiske modeller i gymnasiet og HF?

De stokastiske modeller, vi har set på, er opstået ved at tilpasse dele af den almindelige teori for stokastiske processer til fænomener i et givet genstandsområde. Vi er af den opfattelse, at disse modeller giver et udmærket indblik i, hvad modeller er i almindelighed, og hvad stokastiske modeller er i særdeleshed. Det er imidlertid et stort problem med hensyn til formidling, at teorien for stokastiske processer kræver et ret indgående kendskab til sandsynlighedsregning, f.eks. forskellige sandsynlighedsfordelinger, betingede sandsynligheder, Markov-kæder, simultane fordelinger, middelværdi- og variansberegninger. Desuden kræves kendskab til en række matematiske begreber, som f.eks. iteration, kontraktion, rækkeudvikling, integral- og differentialregning.

Så konklusionen på spørgsmålet må være, at stokastiske modeller af den type, vi har beskæftiget os med, måske nok er relevante, men meget vanskelige at formidle. Den eneste mulighed må være, at lade forgreningsmodellen repræsentere den stokastiske del af et specialestudium i matematiske modeller. Dette ville dog kræve en ret målbevidst faglig indsats, hvilket ville gå ud over de mere generelle aspekter omkring matematiske modeller.

Ser vi bort fra forgreningsmodellen og diffusionsmodellen findes der andre, mere simple dele af populationsgenetikken, der kan anvendes til at eksemplificere begreber fra sandsynlighedsregningen.

Normalt starter sandsynlighedsregningen i gymnasiet med kombinatorik, hvorefter man introducerer sandsynlighedsteorien i overvejende grad eksemplificeret ved kast med mønt og terning, trækning af kugler fra urner o.l. Hermed forsøges at illustrere de tilfældighedsmekanismer, som sandsynlighedsregningen skal behandle. Ved også at anvende populationsgenetiske eksempler til at illustrere sandsynlighedsregningens begreber, får man i højere grad præsenteret matematikkens anvendelse i andre fag i form af matematiske modeller. Man bemærker desuden muligheden for etableringen af et tværfagligt samarbejde mellem biologi og matematik via populationsgenetikken.

Vi vil i det følgende vise, hvorledes nogle af de mest elementære sandsynlighedsbegrebers anvendelse kan illustreres ud fra populationsgenetiske eksempler. Gennemgangen har umiddelbart to svage sider. For det første bliver det nemmere at illustrere modelbegrebet, når matematikken har nået en vis kompleksitet, så vor illustration af modelbegrebet i denne sammenhæng bliver nødvendigvis mangelfuld. For det andet kræver eksemplerne et kendskab til populationsgenetik svarende til kapitel 1 i dette projekt. Eksemplerne egner sig således bedst til en matematisk-biologisk gymnasieklasse eller en HF-klasse med biologi og matematik som tilvalg.

Eksempel, der skal tjene til at afklare begreberne hændelse og udfaldsrum, og anvendelserne af disse:

Vi antager, at 2 forældre med genotypen Aa krydses. Hver af forældrene producerer altså 2 slags kønsceller, A og a, og afkommets genotype er et resultat af en forening af en kønscelle fra hver af forældrene.

Mulige genotyper (udfald) :

♀ \ ♂	A	a
A	AA	Aa
a	aA	aa

Vi har altså et udfaldsrum bestående af 4 udfald.

Genotyperne aA og Aa kan regnes for at være ens. I så fald deles udfaldsrummet kun op i 3 hændelser. Det betyder dog et tab af information om, hvem der har bidraget med hvilke gener til genotypen.

Er vi kun interesseret i fænotyper, deles udfaldsrummet op i 2 hændelser A- og aa. Ved en sådan opdeling af udfaldsrummet mister vi desuden information om nogle af genotypernes art.

Hvor mange hændelser skal man dele udfaldsrummet op i ?

Dette afhænger naturligvis af, hvilket problem vi er interesseret i. Man skal vælge den inddeling af sit udfaldsrum, der giver den simpleste, fyldestgørende forklaring på det problem.

Eksempel, der skal tjene til at problematisere begrebet sandsynlighed som en grænse for den relative frekvens.

Mendel bestemte forholdet mellem dominante og recessive fænotyper hos ærteblomster til at være 3 : 1. Ved krydsningen Aa x Aa, hvor A er genet for rød blomsterfarve, og a er genet for hvid blomsterfarve, ville der komme ialt N planter. N₁ planter ville være røde - fænotypen A -, N₂ planter ville være hvide - fænotypen a. N₁ observeredes altså i forholdet N₁/N₁+N₂. Et sådant forhold kaldes en relativ frekvens. Ved at gentage eksperimentet mange gange bemærkede Mendel, at den relative frekvens nærmede sig en grænse lig 3/4.

En måde at opfatte sandsynlighed på er netop som en grænse for den relative frekvens. Dette kaldes ofte den statistiske definition af sandsynlighed.

Hvilke problemer er der ved at fastlægge en sandsynlighed på denne måde ?

Eksempel, der skal vise, hvorledes regnereglerne for uafhængige begivenheder benyttes i en populationsgenetisk model:

I 1908 generaliserede Hardy og Weinberg uafhængigt af hinanden Mendels love. De benyttede sig af regnereglerne for uafhængige hændelser til at beskrive, hvordan genotyperne vil fordele sig i en population med genfrekvenserne P(A) og P(a).

Resultatet blev følgende:

$$P(AA) = P(A) P(A) = P(A)^2$$

$$P(Aa) = P(A) P(a) + P(a) P(A) = 2 P(A) P(a)$$

$$P(aa) = P(a) P(a) = P(a)^2$$

Denne lov kaldes kort Hardy Weinbergs lov.

Hvilke betingelser skal være opfyldt for at Hardy Weinbergs lov gælder ?

Når Hardy Weinbergs lov benytter sig af regnereglen for uafhængige hændelser, betyder det, at man kun kan være sikker på, at loven gælder, hvor parringen sker tilfældigt. Dette indebærer, at populationer skal opfylde visse biologiske krav, for at man med rimelighed kan benytte loven. Der må f.eks. ikke finde assortativ parring sted - dvs. individer med visse genotyper må ikke at parre sig med individer med bestemte genotyper -, ligeledes må der ikke være selektion, migration m.v. Havde man indregnet disse mange forhold i loven/modellen, havde genstandsområdet kompleksitet gjort det meget vanskeligt at opstille et korrekt udtryk for genotypefrekvenserne. Ved at holde sig til det simple tilfælde, hvor den tilhørende teori er overskuelig og velunderbygget, bliver resultatet meget troværdigt. Til gengæld lægger man så mange bånd på de populationer, loven beskriver, at man kun yderst sjældent vil finde populationer i naturen, der opfylder alle betingelserne.

Eksempel, der skal illustrere forskellige fordelinger og vanskelighederne ved at vælge mellem forskellige fordelinger i en konkret situation:

Indenfor populationsgenetik benyttes forskellige afkomstfordelinger for f.eks. mutante gener.

Binomialfordeling :
$$P_k = \binom{2N}{k} \cdot \left(\frac{1}{2N}\right)^k \left(1 - \frac{1}{2N}\right)^{2N-k}$$

Dette er sandsynligheden for, at et mutant gen i generation t bliver til k mutante gener i generation t+1 i en population med størrelsen N.

Poissonfordeling :
$$P_k = \frac{M^k}{k!} \cdot e^{-M}$$

Dette er sandsynligheden for, at et mutant gen i generation t bliver til k mutante gener i generation t+1, når middeltal afkom for det mutante gen er lig M.

Diskussionsspørgsmål:

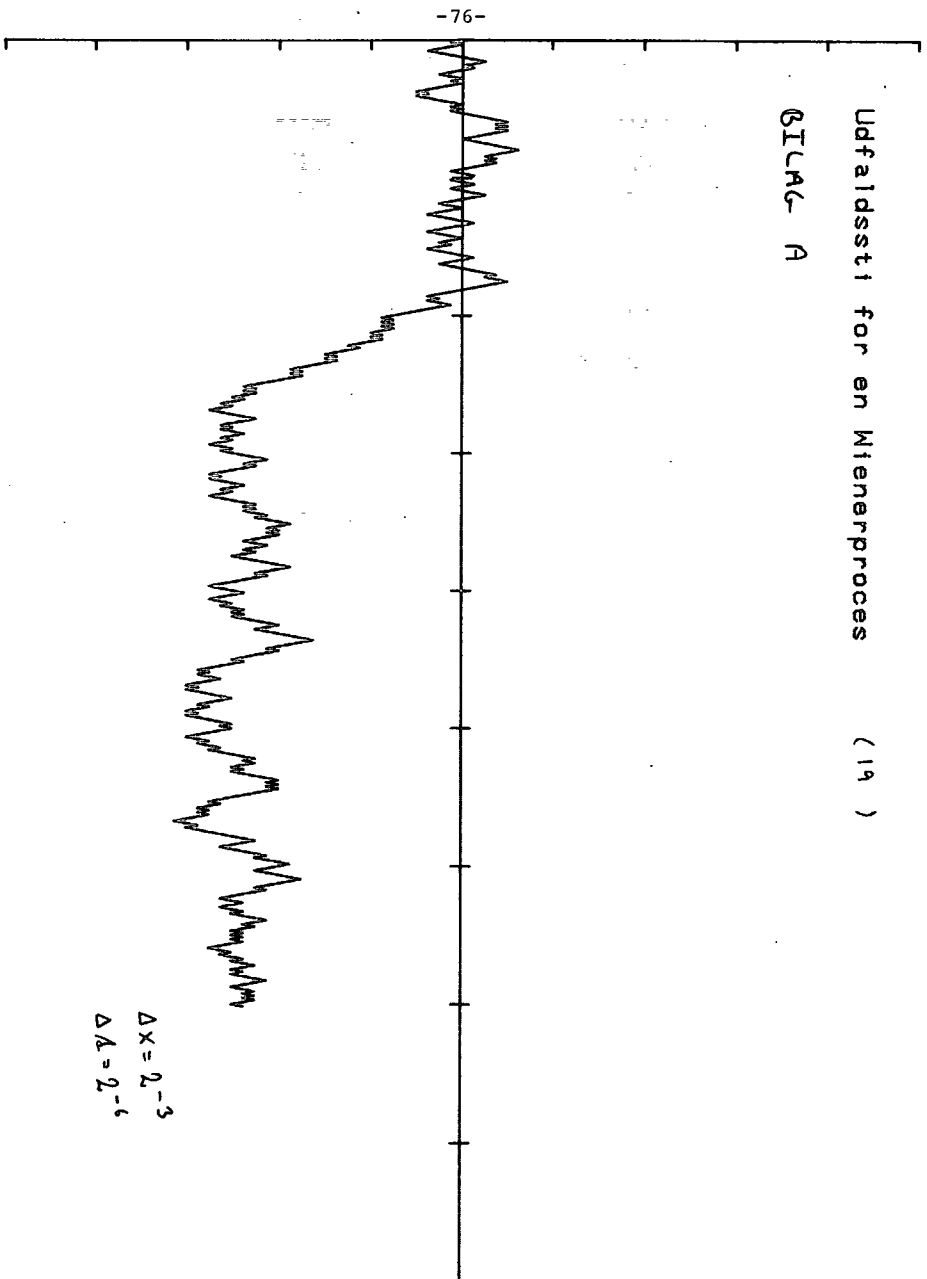
Hvornår er det bedre at anvende en binomialfordeling end en poissonfordeling ?

Alt i alt må man sige, at der er meget få begreber i sandsynlighedsregningen, som ikke kan illustreres med populationsgenetiske eksempler. Ud fra en pædagogisk/matematisk synsvinkel kan man sige, at begrænsningen mest ligger i, hvor meget tid man ønsker at anvende til at forklare begreber fra populationsgenetikken.

Udfaldssti for en Wienerproces (19)

BILAG A

$$\begin{aligned}\Delta x &= 2^{-3} \\ \Delta t &= 2^{-6}\end{aligned}$$

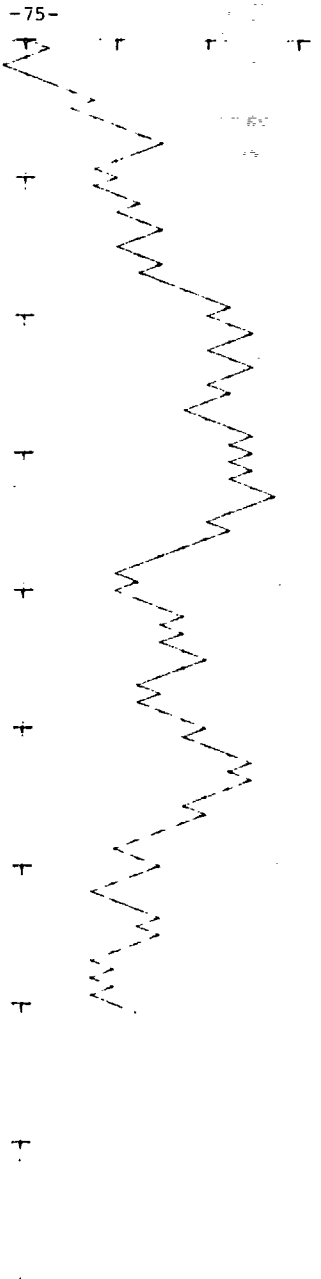


-76-

Udfaldssti for en Wienerproces (19)

BILAG A

$$\begin{aligned}\Delta x &= \frac{1}{4} \\ \Delta t &= \frac{1}{16}\end{aligned}$$



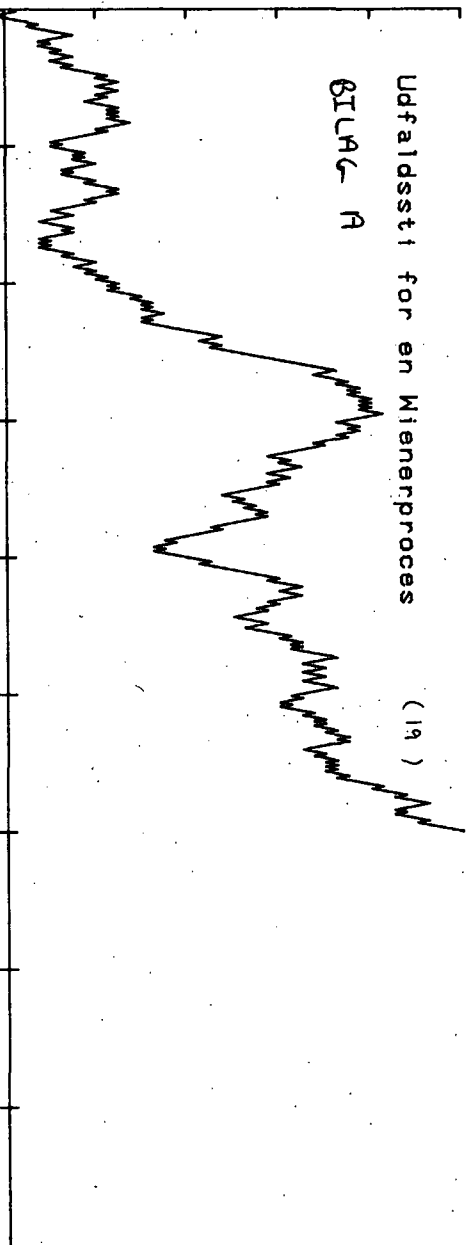
-75-

Udfaldssti for en Wienerproces (19)
BILAG A



$$\Delta t = 2^{-10}$$

Udfaldssti for en Wienerproces (19)
BILAG A

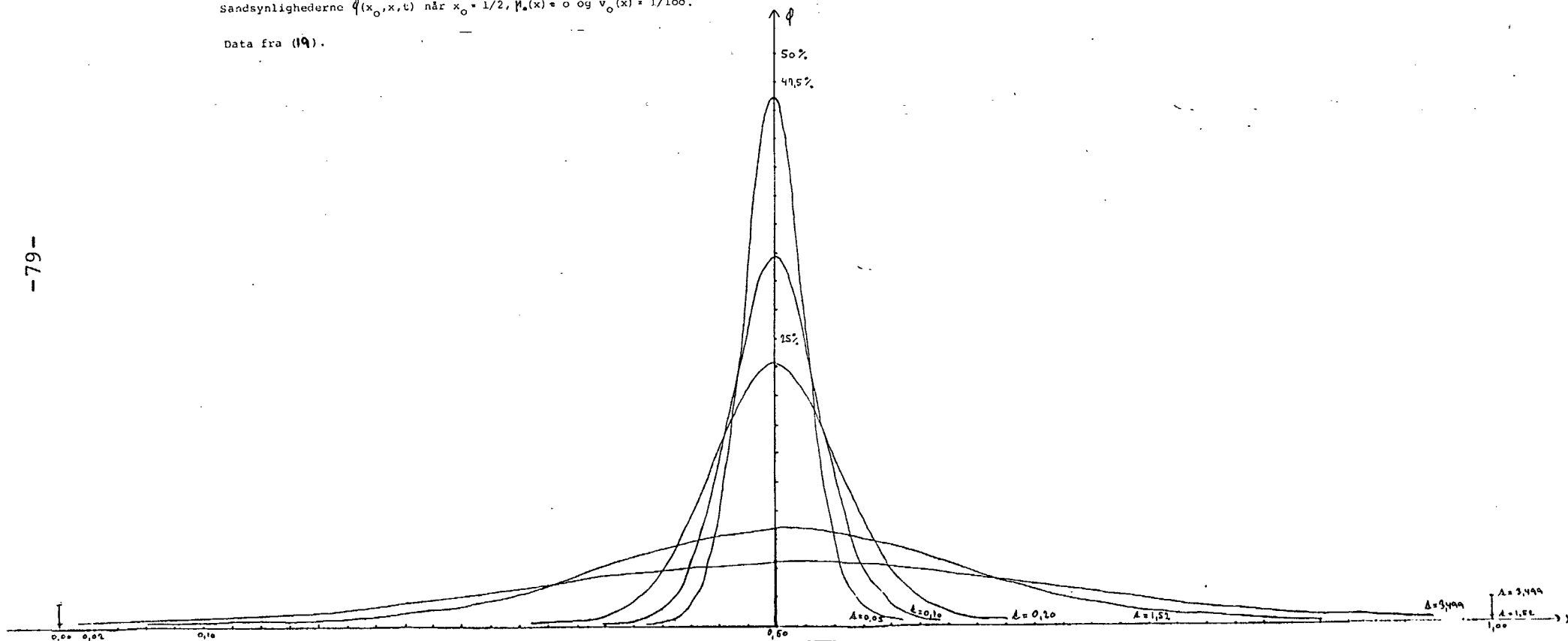


$$\Delta X = 2^{-3}$$
$$\Delta t = 2^{-6}$$

BILAG B.

Sandsynlighederne $\phi(x_0, x, t)$ når $x_0 = 1/2$, $\mu_0(x) = 0$ og $v_0(x) = 1/100$.

Data fra (19).



BILAG C. Middelværdi- og variansberegninger.

Vi vil først beregne middelværdien af ændringen i genfrekvensen i det tilfælde, hvor der ingen selektion optræder ($s = 0$). Vi skal altså finde $E(Y/2N_e - x_0)$, hvor $Y/2N_e$ er en stokastisk varierende genfrekvens i generation 1, og x_0 er den konstante startfrekvens i generation 0. Vi har da, at $E(Y/2N_e - x_0) = E(Y/2N_e) - x_0$. Eftersom der ingen selektion er, dvs. frekvensen ikke systematisk presses i nogen retning, antages at $E(Y/2N_e) = E(x_0) = x_0$. Altså er $E(Y/2N_e - x_0) = x_0 - x_0 = 0$.

Dernæst ser vi på variansen i frekvensændringen. I tilfældet med $s=0$ har vi, idet Y betegner antallet af mutante gener i generation 1, at $Y = B_1 + \dots + B_{2N_e}$, hvor B_1, \dots, B_{2N_e} er uafhængige bernoulli variable (hvilket er det samme som at Y er binomialfordelt). Da er

$$V\left(\frac{Y}{2N_e}\right) = \frac{V(Y)}{(2N_e)^2} = \frac{V(B_1 + \dots + B_{2N_e})}{(2N_e)^2} = \frac{2N_e x(1-x)}{(2N_e)^2} = \frac{x(1-x)}{2N_e}$$

Og

$$V_0(x - x_0) = V(x) - V(x_0) = V(x) = \frac{x(1-x)}{2N_e}$$

Bemærk at vi her indfører N_e i stedet for N - N_e er den såkaldte effektive populationsstørrelse. En population med N individer formerer sig ikke nødvendigvis særligt effektivt - N_e kan opfattes som det mindste antal individer, der kan formere sig lige så effektivt. N_e er altså en modelpopulation, der formerer sig med optimal effektivitet.

Hvis f.eks. N_m er antallet af hanner, og N_f er antallet af hunner i en population - altså $N = N_m + N_f$ - kan N_e vises at være $N_e = 4N_m N_f / N$ (9). Heraf haves, at $N_e = N$ når $N_m = N_f$, og at $N_e < N$ når $N_m \neq N_f$. Andre ting af betydning for forholdet mellem N_e og N kan være geografiske forhold (9).

Vi skal så til at se på tilfælde med selektion, hvor middelen- dringen i genfrekvensen bliver forskellig fra 0. Her vil middel- ændringen være givet ved forskellige udtryk afhængigt af, hvilke former for selektion, der er tale om. Derimod vil udtrykket for variansen forblive uændret, da der blot er tale om, at processen systematisk presses i en bestemt retning med en konstant s -værdi forskellig fra 0, hvilket ikke giver anledning til ændret varians.

Vi vil ikke her give noget formaliseret bevis for denne på- stand. Men vi har altså, at for to forskellige former for se- lektion, som vi skal beregne middelændringen i genfrekvensen for i det følgende er $V_0(x_0) = x(1-x)/2N_e$.

BILAG C. Middelværdi- og variansberegninger.

Vi ser først på tilfældet, hvor A_1 har en konstant selektiv fordel i forhold til A_2 (6):

	A_1A_1	A_1A_2	A_2A_2
	1+2s	1+s	1
generation t:	x^2	$2x(1-x)$	$(1-x)^2$ (genotypefre-
			kvenser, når A_1 har
			frekvensen x).

Dette svarer til et forhold mellem genotypefrekvenser på $x^2(1+2s)/2x(1-x)(1+s)/(1-x)^2$ i generation t+1, når selektionen indregnes. Idet x' betegner middelgenfrekvensen i generation t+1 og x den tilsvarende i generation t, fås middelændringen i frekvensen til:

$$x' - x = \frac{x^2(1+2s) + x(1-x)(1+s)}{x^2(1+2s) + 2x(1-x)(1+s) + (1-x)^2} - x$$

$$= \frac{2sx^2 + x + sx(1-x)}{1 + 2sx^2 + 2sx(1-x)} - x$$

$$= \frac{2sx^2 + sx(1-x) - 2sx^3 - 2sx^2(1-x)}{1 + 2sx^2 + 2sx(1-x)}$$

$$= \frac{sx(1-x)}{1 + (2x^2 + 2x(1-x))s}$$

$$f(s) = \frac{sx(1-x)}{1 + (2x^2 + 2x(1-x))s} = \frac{as}{1+bs}, \text{ hvor } a = x(1-x) \text{ og } b = 2x^2 + 2x(1-x)$$

$$f(0) = 0; f'(s) = a \cdot \frac{(1+bs) - s \cdot b}{(1+bs)^2} = \frac{a}{(1+bs)^2}$$

$$f''(0) = a - x(1-x); f''(s) = a \cdot \frac{-2(1+bs) \cdot b}{(1+bs)^4} = \frac{-2b}{(1+bs)^3}$$

$$f'''(0) = -2b = -4x^2 - 4x(1-x) = -4x$$

Vi har da at:

$$f(s) = x(1-x)s + (-2x)s^2 + \dots$$

og for lille |s| høves: $f(s) \approx sx(1-x)$

BILAG C. Middelværdi- og variansberegninger.

Ser vi herefter på et mere generelt tilfælde, hvor s og $s \cdot h$ er hhv. den selektive fordel for den mutante homozygot og heterozygoten, har vi:

	A_1A_1	A_1A_2	A_2A_2
	1+s	1+hs	1

Og middelændringen i frekvensen bliver da:

$$x' - x = \frac{x^2(1+s) + x(1-x)(1+hs)}{x^2(1+s) + 2x(1-x)(1+hs) + (1-x)^2} - x$$

$$= \frac{x^2s + x + x(1-x)hs}{1 + sx^2 + 2x(1-x)hs} - x$$

$$= \frac{x^2s + x(1-x)hs - sx^3 - 2x^2(1-x)hs}{1 + s(x^2 + 2x(1-x))}$$

$$= \frac{sx(1-x)(x+h-2xh)}{1 + s(x^2 + 2x(1-x))}$$

$$= \frac{sx(1-x)(h + (1-2h)x)}{1 + s(x^2 + 2x(1-x))}$$

$$f(s) = \frac{as}{1+bs} \text{ rækkeudvikles omkring } s=0:$$

$$f(0) = 0; f'(s) = a \cdot \frac{(1+bs) - s \cdot b}{(1+bs)^2} = a \cdot \frac{1}{(1+bs)^2}; f'(0) = a$$

$$f''(s) = a \cdot \frac{2(1+bs) \cdot b}{(1+bs)^4} = \frac{2ab}{(1+bs)^3}; f''(0) = 2ab$$

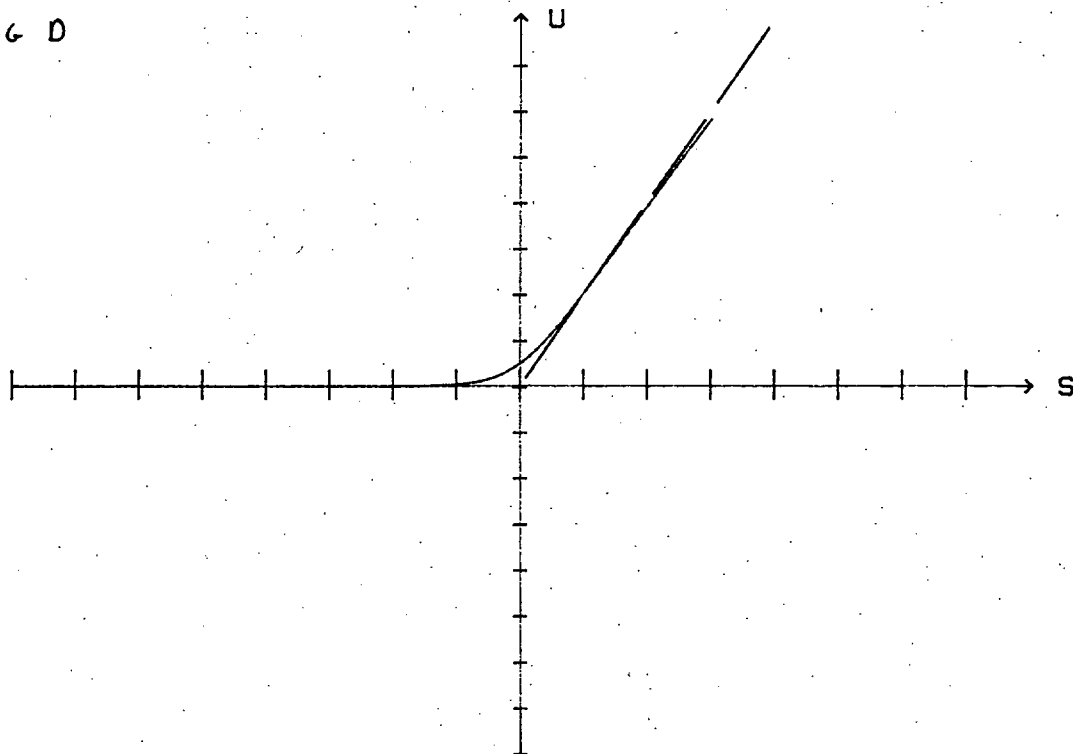
$$f(s) = sx(1-x)(x+h-2xh) + \frac{s^2}{2} (x^2 + 2x(1-x))2x(1-x)(x+h-2xh) + \dots$$

for |s| lille høves

$$f(s) \approx sx(1-x)(h + (1-2h)x)$$

FIGUR 3.4.2.B $U = (1 - \exp(-2*S)) / (1 - \exp(-4*N*S))$; $NE=N=100$; $-0.08 \leq S \leq 0.03$
 $U = 2*S$; $0 < S \leq 0.04$; STIPLET

BILAG D

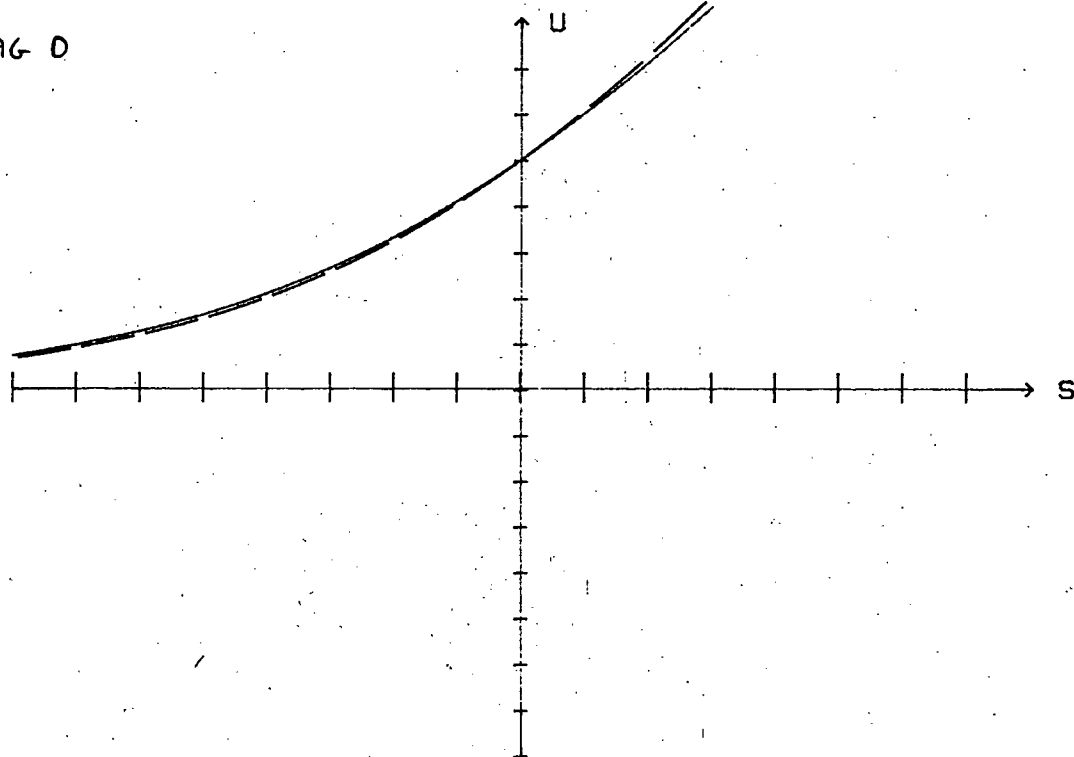


-85-

DEN STIPLEDE GRAF SKAL GIVE EN GOD APPROXIMATION FOR $S > 0$ OG N STOR

FIGUR 3.4.2.A $U = (1 - \exp(-2*S)) / (1 - \exp(-4*N*S))$; $NE=N=10$; $-0.08 \leq S \leq 0.03$
 $U = 2*S / (1 - \exp(-4*N*S))$; $NE=N=10$; $-0.08 \leq S \leq 0.03$; STIPLET

BILAG D



-84-

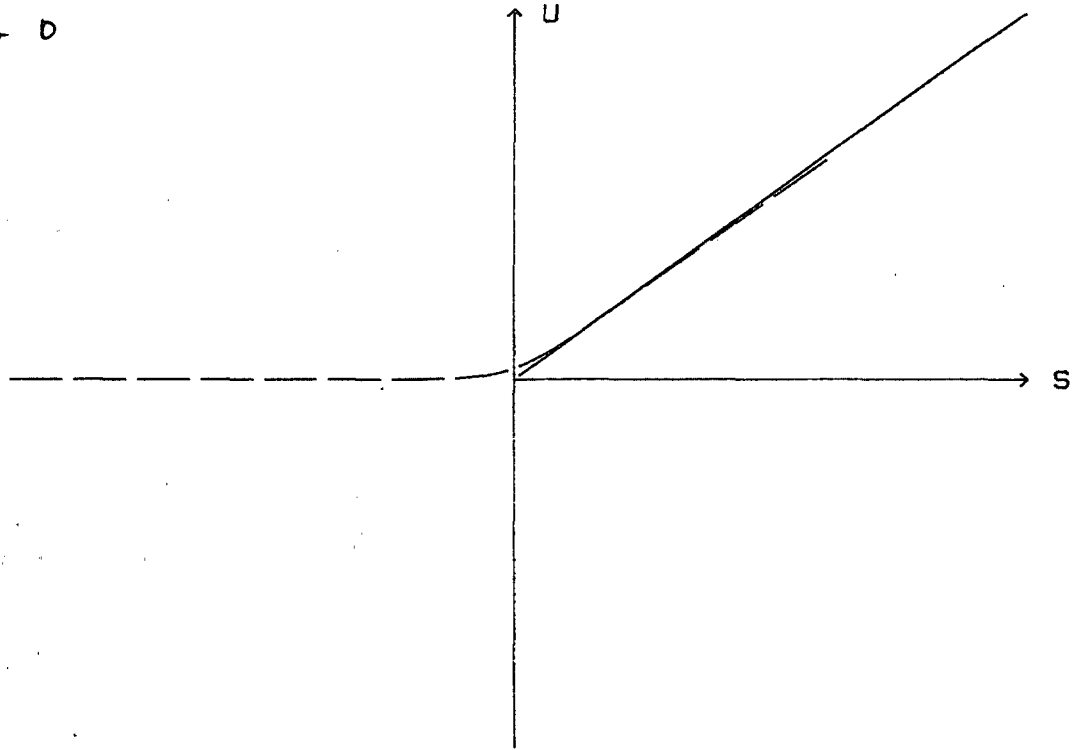
DEN STIPLEDE GRAF SKAL GIVE EN GOD APPROXIMATION FOR MEGET LILLE ABS(S)

FIGUR 3.4.2.D

BILAG D

$$U = (1 - \exp(-2 * NE * S / N)) / (1 - \exp(-4 * NE * S)); NE = 100; N = 200; \text{STIPLET}$$
$$U = 2 * NE * S / N; NE = 100; N = 200$$

-87-



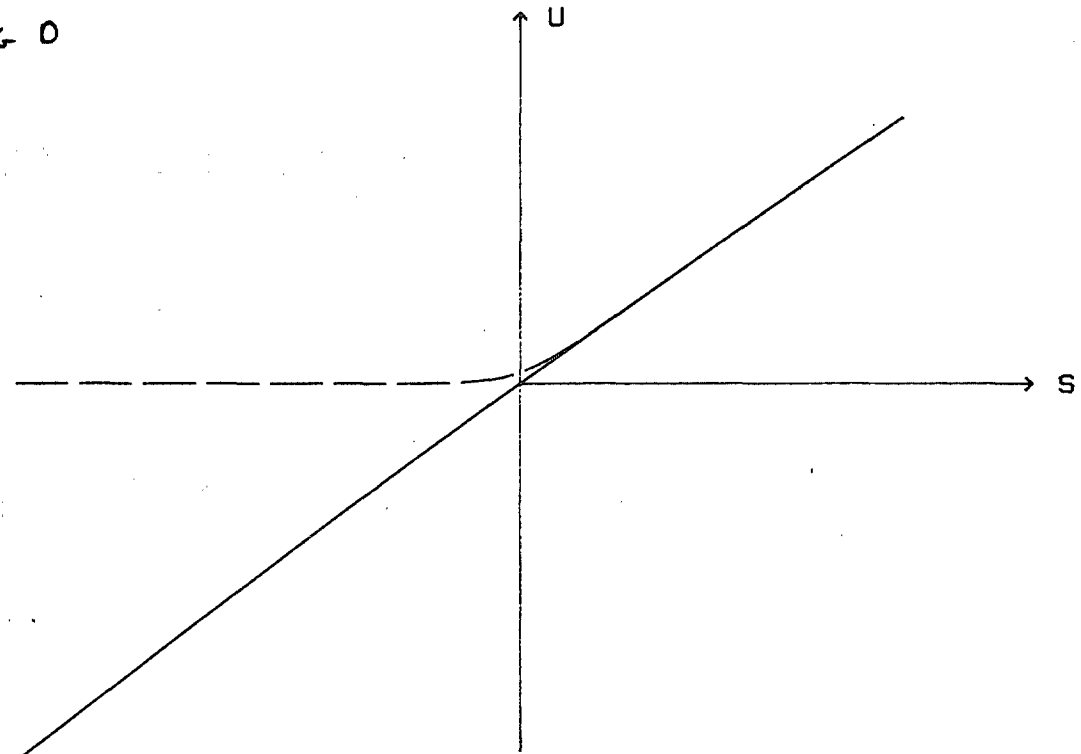
DEN FULDT OPTRUKNE GRAF SKAL GIVE EN GOD APPROXIMATION FOR $4NE * S \gg 1$ OG S LILLE OG POSITIV

FIGUR 3.4.2.C

BILAG D

$$U = (1 - \exp(-2 * NE * S / N)) / (1 - \exp(-4 * NE * S)); NE = 100; N = 200; \text{STIPLET}$$
$$U = 1 - \exp(-2 * NE * S / N); NE = 100; N = 200$$

-86-



DEN FULDT OPTRUKNE GRAF ER EN GOD APPROXIMATION FOR $4NE * S \gg 1$, DVS. S POS. OG NE STOR

BILAG E. Omskrivning af udtrykket for overlevelsessandsynligheden.

Vi ønsker at få udtrykket

$$u(x_0) = \frac{e^{-4Nesx_0} - 1}{e^{-4Nes} - 1}$$

skrevet op som en række. I en matematisk opslagsbog finder man, at Bernoullis polynomier B_n er defineret ved:

$$t \cdot \frac{e^{zt}}{e^t - 1} = \sum_{n=1}^{\infty} B_n(z) t^n$$

$$\Leftrightarrow \frac{e^{zt}}{e^t - 1} = \sum_{n=1}^{\infty} B_n(z) t^{n-1}$$

hvor $|t| < 2\pi$ og de første tre Bernoulli-polynomier er

$$B_1(z) = z - \frac{1}{2}; B_2(z) = (z^2/2) - (z/2) + \frac{1}{12}; B_3(z) = (z^3/6) - (z^2/4) + (z/2)$$

Vi har nu

$$\frac{e^{zt} - 1}{e^t - 1} = -\frac{1}{e^t - 1} + \frac{e^{zt}}{e^t - 1} = -\sum_{n=1}^{\infty} B_n(0) t^{n-1} + \sum_{n=1}^{\infty} B_n(z) t^{n-1}$$

$$= \sum_{n=1}^{\infty} (B_n(z) t^{n-1} - B_n(0) t^{n-1})$$

$$= \sum_{n=1}^{\infty} (B_n(z) - B_n(0)) t^{n-1} \quad \text{for } t = -4Nes \text{ og } z = x_0$$

Heraf fåes

$$u(x_0) = \sum_{n=1}^{\infty} (B_n(x_0) - B_n(0)) \cdot (-4Nes)^{n-1}$$

Vi vil beregne de tre første led

$n = 1$:

$$(B_1(x_0) - B_1(0)) \cdot 1 = x_0 - \frac{1}{2} - (-\frac{1}{2}) = \underline{x_0}$$

$n = 2$:

$$(B_2(x_0) - B_2(0)) \cdot (-4Nes)$$

$$= ((\frac{x_0^2}{2} - \frac{x_0}{2} + \frac{1}{12}) - (\frac{1}{12})) \cdot (-4Nes)$$

$$= (-2Nes) \cdot x_0 \cdot (1 - x_0)$$

$$= \underline{\underline{2Nes x_0 (1 - x_0)}}$$

$n = 3$:

$$(B_3(x_0) - B_3(0)) \cdot (-4Nes)^2$$

$$= ((\frac{x_0^3}{6} - \frac{x_0^2}{4} + \frac{x_0}{2} - (0)) \cdot (-4Nes)^2$$

$$= (\frac{2}{3} x_0^3 - x_0^2 + \frac{1}{3} x_0) \cdot (2Nes)^2$$

$$= \frac{(2Nes)^2}{3} \cdot x_0 (2x_0^2 - 3x_0 + 1)$$

$$= \underline{\underline{\frac{(2Nes)^2}{3} \cdot x_0 (x_0 - 1) (2x_0 - 1)}}$$

Vi skriver altså nu $u(x_0)$ som

$$|2Nes| < \pi :$$

$$u(x_0) = x_0 + 2Nes x_0 (1 - x_0) + \frac{(2Nes)^2}{3} \cdot x_0 (x_0 - 1) (2x_0 - 1) + \dots$$

LITTERATURLISTE

1. Bailey, N.T.J. (1977). "Mathematics, Statistics, and Systems for health" Wiley, The Pitmann Press, Bath.
2. Bartlett, M.S. (1966). "An Introduction to Stochastic Processes", 2.nd ed., Cambridge University Press, London.
3. Bulmer, M.G. (1980). "The mathematical theory of Quantitative Genetics", Clarendon Press, Oxford.
4. Crow, J.F. and Kimura, M. (1970). "An Introduction to Population Genetics Theory", Harper and Row, New York.
5. Crow, J.F. (1972). "Darwinian and Non-darwinian Evolution", Proceedings of the Sixth Berkeley Symposium, Vol. V: 43-69.
6. Falconer, D.S. (1981). "Introduction to Quantitative Genetics", 2.nd ed., Longman, New York.
7. Fisher, R.A. (1930). "The Genetical Theory og Natural Selection", Clarendon Press, Oxford.
8. Haldane, J.B.S. (1927). "A mathematical theory of natural and artificial selection", Part V. Selection and mutation. Proc. Camb. Phil. Soc. 23: 838-844.
9. Hartl, D.L. (1981). "A Primer of Population Genetics", Sinauer, Massachusetts.
10. Iyanaga, S. and Kawada, Y. (1977). "Encyclopedic Dictionary of Mathematics", vol. 1, oversat af May, K.D., Massachusetts.
11. Jensen, J.H. (1980). "Matematiske modeller - vejledning eller vildledning", Naturkampen 18.
12. Kendall, D.G. (1966). "Branching Processes since 1873", Journal London Math. Soc. 41: 385-406.
13. Kimura, M. and Ohta, T. (1971). "Theoretical Aspects of Population Genetics", Princeton Univ., Princeton.
14. Kimura, M. and Ohta, T. (1972). "Population Genetics, Molecular Biometry, and Evolution", Proceedings of the Berkeley Symposium, Vol. V: 43-69.
15. King, J.L. (1972). "The Role of Mutation in Evolution", Proceedings of the Sixth Berkeley Symposium, Vol. V: 69-98.
16. Kojima, K. and Kelleher, T.M. (1962). "Survival of mutant genes", Amer. Nat. 96: 329-346.
17. Krista-gruppen. (1981). "Hvor ligger Ballerup", Naturkampen 21.
18. Lange, T. (1982). "En - to - mange. En undersøgelse af matematisk økologi", IMFUFA-tekst nr. 56.
19. Larsen, Jørgen (1983)
20. Livestock, J. (1974). "The New Encyclopædia Britannica", Vol. 10., University of Chicago. s. 501-503.
21. Witt-Hansen, J. (1963). "Om generalisation og generalisationsproblemer i de matematiske og historiske videnskaber", Festskrift udgivet af Københavns Universitet i anledning af universitetets festskrift.

- 1/78 "TANKER OM EN PRAKSIS" - et matematikprojekt.
Projekt rapport af Anne Jensen, Lena Lindenskov, Marianne Kesselhahn og Nicolai Lomholt.
Vejleder: Anders Madsen.
- 2/78 "OPTIMERING" - Menneskets forøgede beherskelsesmuligheder af natur og samfund.
Projekt rapport af Tom J. Andersen, Tommy R. Andersen, Gert Kreinøe og Peter H. Lassen.
Vejleder: Bernhelm Booss.
- 3/78 "OPGAVESAMLING", breddekursus i fysik.
Lasse Rasmussen, Aage Bonde Kræmmer, Jens Højgaard Jensen.
- 4/78 "TRE ESSAYS" - om matematikundervisning, matematiklæreruddannelsen og videnskabsrindalismen. Nr. 4 er p.t. udgået.
Mogens Niss.
- 5/78 "BIBLIOGRAFISK VEJLEDNING til studiet af DEN MØDERNE FYSIKS HISTORIE".
Helge Kragh.
- 6/78 "NOGLE ARTIKLER OG DEBATINDLÆG OM - læreruddannelse og undervisning i fysik, og - de naturvidenskabelige fags situation efter studentereprøret".
Karin Beyer, Jens Højgaard Jensen og Bent C. Jørgensen.
- 7/78 "MATEMATIKKENS FORHOLD TIL SAMFUNDSØKONOMIEN". Nr. 7 er udgået.
B.V. Gnedenko.
- 8/78 "DYNAMIK OG DIAGRAMMER". Introduktion til energy-bound-graph formalismen.
Peder Voetmann Christiansen.
- 9/78 "OM PRAKSIS' INDFLYDELSE PÅ MATEMATIKKENS UDVIKLING": - Motiver til Kepler's: "Nova Stereometria Doliorum Vinariorum".
Projekt rapport af Lasse Rasmussen.
Vejleder: Anders Madsen.
-
- 10/79 "TERMODYNAMIK I GYMNASIET".
Projekt rapport af Jan Christensen og Jeanne Mortensen.
Vejledere: Karin Beyer og Peder Voetmann Christiansen.
- 11/79 "STATISTISKE MATERIALER"
red. Jørgen Larsen
- 12/79 "LINEÆRE DIFFERENTIALLIGNINGER OG DIFFERENTIALLIGNINGSSYSTEMER".
Mogens Brun Heefelt
- 13/79 "CAVENDISH'S FORSØG I GYMNASIET".
Projekt rapport af Gert Kreinøe.
Vejleder: Albert Chr. Paulsen

- 14/79 "BOOKS ABOUT MATHEMATICS: History, Philosophy, Education, Models, System Theory, and Works of Reference etc. A Bibliography".
Else Høyrup. Nr. 14 er p.t. udgået.
- 15/79 "STRUKTUREL STABILITET OG KATASTROFER i systemer i og udenfor termodynamisk ligevægt".
Specialeopgave af Leif S. Striegler.
Vejleder: Peder Voetmann Christiansen.
- 16/79 "STATISTIK I KRÆFTFORSKNINGEN".
Projektrapport af Michael Olsen og Jørn Jensen.
Vejleder: Jørgen Larsen.
- 17/79 "AT SPØRGE OG AT SVARE i fysikundervisningen".
Albert Christian Paulsen.
- 18/79 "MATHEMATICS AND THE REAL WORLD", Proceedings of an International Workshop, Roskilde University Centre, Denmark, 1978. Preprint.
Bernhelm Booss & Mogens Niss (eds.).
- 19/79 "GEOMETRI, SKOLE OG VIRKELIGHED".
Projektrapport af Tom J. Andersen, Tommy R. Andersen og Per H.H. Larsen.
Vejleder: Mogens Niss.
- 20/79 "STATISTISKE MODELLER TIL BESTEMMELSE AF SIKRE DOSER FOR CARCINOGENE STOFFER".
Projektrapport af Michael Olsen og Jørn Jensen.
Vejleder: Jørgen Larsen.
- 21/79 "KONTROL I GYMNASIET - FORMAL OG KONSEKVENSER".
Projektrapport af Crilles Bacher, Per S. Jensen, Preben Jensen og Torben Nysteen.
- 22/79 "SEMIOTIK OG SYSTEMEGENSKABER (1)".
1-port lineært response og støj i fysikken.
Peder Voetmann Christiansen.
- 23/79 "ON THE HISTORY OF EARLY WAVE MECHANICS - with special emphasis on the role of reality".
-
- 24/80 "MATEMATIKOPFATTELSE HOS 2.G'ERE".
a+b 1. En analyse. 2. Interviewmateriale.
Projektrapport af Jan Christensen og Knud Lindhardt Rasmussen. Nr. 24 a+b er p.t. udgået.
Vejleder: Mogens Niss.
- 25/80 "EKSAMENSOPGAVER", Dybdemodulet/fysik 1974-79.
- 26/80 "OM MATEMATISKE MODELLER".
En projektrapport og to artikler. Nr. 26 er p.t. udgået.
Jens Højgaard Jensen m.fl.
- 27/80 "METHODOLOGY AND PHILOSOPHY OF SCIENCE IN PAUL DIRAC'S PHYSICS".
Helge Kragh.
- 28/80 "DIELEKTRISK RELAXATION - et forslag til en ny model bygget på væskernes viscoelastiske egenskaber".
Projektrapport, speciale i fysik, af Gert Kreinøe.
Vejleder: Niels Boye Olsen.

- 29/80 "ODIN - undervisningsmateriale til et kursus i differentialligningsmodeller".
Projektrapport af Tommy R. Andersen, Per H.H. Larsen og Peter H. Lassen.
Vejleder: Mogens Brun Heefelt
- 30/80 "FUSIONSENERGIEN - - - ATOMSAMFUNDETS ENDESTATION".
Oluf Danielsen.
Nr. 30 er udgået.
Udkommer medio 1982 på Fysik-, Matematik- og Kemilærer-nes forlag.
- 31/80 "VIDENSKABSTEORETISKE PROBLEMER VED UNDERVISNINGSSYSTEMER BASERET PÅ MÆNGDELÆRE".
Projektrapport af Troels Lange og Jørgen Karrebæk.
Vejleder: Stig Andur Pedersen.
Nr. 31 er p.t. udgået
- 32/80 "POLYMERE STOFFERS VISCOELASTISKE EGENSKABER - BELYST VED HJÆLP AF MEKANISKE IMPEDANSMÅLINGER OG MOSSBAUER-EFFEKTMÅLINGER".
Projektrapport, speciale i fysik, af Crilles Bacher og Preben Jensen.
Vejledere: Niels Boye Olsen og Peder Voetmann Christiansen.
- 33/80 "KONSTITUERING AF FAG INDEN FOR TEKNISK-NATURVIDENSKABELIGE UDDANNELSER. I-II".
Arne Jakobsen.
- 34/80 "ENVIRONMENTAL IMPACT OF WIND ENERGY UTILIZATION".
ENERGY SERIES NO.1.
Bent Sørensen.
Nr. 34 er udgået.
Publ. i "Renewable Sources of Energy and the Environment", Tycooli International Press, Dublin, 1981.
- 35/80 "HISTORISKE STUDIER I DEN NYERE ATOMFYSIKS UDVIKLING".
Helge Kragh.
- 36/80 "HVAD ER MENINGEN MED MATEMATIKUNDERVISNINGEN ?".
Fire artikler.
Mogens Niss.
- 37/80 "RENEWABLE ENERGY AND ENERGY STORAGE".
ENERGY SERIES NO.2.
Bent Sørensen.
-
- 38/81 "TIL EN HISTORIE TEORI OM NATURERKENDELSE, TEKNOLOGI OG SAMFUND".
Projektrapport af Erik Gade, Hans Hedal, Henrik Lau og Finn Physant.
Vejledere: Stig Andur Pedersen, Helge Kragh og Ib Thiersen.
- 39/81 "TIL KRITIKKEN AF VÆKSTØKONOMIEN".
Jens Højgaard Jensen.
- 40/81 "TELEKOMMUNIKATION I DANMARK - oplæg til en teknologivurdering".
Projektrapport af Arne Jørgensen, Bruno Petersen og Jan Vedde.
Vejleder: Per Nørgaard.
- 41/81 "PLANNING AND POLICY CONSIDERATIONS RELATED TO THE INTRODUCTION OF RENEWABLE ENERGY SOURCES INTO ENERGY SUPPLY SYSTEMS".
ENERGY SERIES NO.3.
Bent Sørensen.

- 42/81 "VIDENSKAB TEORI SAMFUND - En introduktion til materialistiske videnskabsopfattelser".
Helge Kragh og Stig Andur Pedersen.
- 43/81 1. "COMPARATIVE RISK ASSESSMENT OF TOTAL ENERGY SYSTEMS".
2. "ADVANTAGES AND DISADVANTAGES OF DECENTRALIZATION".
ENERGY SERIES NO.4.
Bent Sørensen.
- 44/81 "HISTORISK UNDERSØGELSE AF DE EKSPERIMENTELLE FORUDSÆTNINGER FOR RUTHERFORDS ATOMMODEL".
Projektrapport af Niels Thor Nielsen.
Vejleder: Bent C. Jørgensen.
-
- 45/82
- 46/82 "EKSEMPLARISK UNDERVISNING OG FYSISK ERKENDELSE - I+II ILLUSTRERET VED TO EKSEMPLER".
Projektrapport af Torben O. Olsen, Lasse Rasmussen og Niels Dreyer Sørensen.
Vejleder: Bent C. Jørgensen.
- 47/82 "BARSEBACK OG DET VÆRST OFFICIELT-TÆNKELIGE UHELD".
ENERGY SERIES NO.5.
Bent Sørensen.
- 48/82 "EN UNDERSØGELSE AF MATEMATIKUNDERVISNINGEN PÅ ADGANGSKURSUS TIL KØBENHAVNS TEKNIKUM".
Projektrapport af Lis Eilertzen, Jørgen Karrebæk, Troels Lange, Preben Nørregaard, Lissi Pedersen, Laust Rishøj, Lill Røn, Isac Showiki.
Vejleder: Mogens Niss.
- 49/82 "ANALYSE AF MULTISPEKTRALE SATELLITBILLEDER".
Projektrapport af Preben Nørregaard.
Vejledere: Jørgen Larsen & Rasmus Ole Rasmussen.
- 50/82 "HERSLEV - MULIGHEDER FOR VEDVARENDE ENERGI I EN LANDSBY". ENERGY SERIES NO.6.
Rapport af Bent Christensen, Bent Hove Jensen, Dennis B. Møller, Bjarne Laursen, Bjarne Lillethorup og Jacob Mørch Pedersen.
Vejleder: Bent Sørensen.
- 51/82 "HVAD KAN DER GØRES FOR AT AFHJÆLPE PIGERS BLOKERING OVERFOR MATEMATIK?"
Projektrapport af Lis Eilertzen, Lissi Pedersen, Lill Røn og Susanne Stender.
- 52/82 "DESUSPENSION OF SPLITTING ELLIPTIC SYMBOLS"
Bernhelm Booss & Krzysztof Wojciechowski.
- 53/82 "THE CONSTITUTION OF SUBJECTS IN ENGINEERING EDUCATION".
Arne Jakobsen & Stig Andur Pedersen.
- 54/82 "FUTURES RESEARCH" - A Philosophical Analysis of Its Subject-Matter and Methods.
Stig Andur Pedersen & Johannes Witt-Hansen.

- 55/82 "MATEMATISKE MODELLER" - Litteratur på Roskilde
Universitetsbibliotek.
En bibliografi.
Else Høyrup.
- Vedr. tekst nr. 55/82:
Se også tekst 62/83.
- 56/82 "ÉN - TO - MANGE" -
En undersøgelse af matematisk økologi.
Projektrapport af Troels Lange.
Vejleder: Anders Madsen.
-
- 57/83 "ASPECT EKSPERIMENTET" -
Skjulte variable i kvantemekanikken?
Projektrapport af Tom Juul Andersen.
Vejleder: Peder Voetmann Christiansen.
- Nr. 57 er udgået.
- 58/83 "MATEMATISKE VANDRINGER" - Modelbetragtninger
over spredning af dyr mellem småbiotoper i
agerlandet.
Projektrapport af Per Hammershøj Jensen &
Lene Vagn Rasmussen.
Vejleder: Jørgen Larsen.
- 59/83 "THE METHODOLOGY OF ENERGY PLANNING".
ENERGY SERIES NO. 7.
Bent Sørensen.
- 60/83 "MATEMATISK MODEKSPERTISE" - et eksempel.
Projektrapport af Erik O. Gade, Jørgen Karrebæk og
Preben Nørregaard.
Vejleder: Anders Madsen.
- 61/83 "FYSIKS IDEOLOGISKE FUNKTION", som et eksempel på
en naturvidenskab - historisk set.
Projektrapport af Annette Post Nielsen.
Vejledere: Jens Høyrup, Jens Højgaard Jensen og
Jørgen Vogelius.
- 62/83 "MATEMATISKE MODELLER" - Litteratur på Roskilde
Universitetsbibliotek.
En bibliografi. 2. rev. udgave
Else Høyrup
- 63/83 "CREATING ENERGY FUTURES: A SHORT GUIDE TO
ENERGY PLANNING".
ENERGY SERIES No. 8
David Crossley & Bent Sørensen
- 64/83 "VON MATHEMATIK UND KRIEG".
Bernhelm Booss og Jens Høyrup
- 65/83 "ANVENDT MATEMATIK - TEORI ELLER PRAKSIS".
Projektrapport af Per Hedegård Andersen, Kirsten
Habekost, Carsten Holst-Jensen, Annelise von Moos,
Else Marie Pedersen, Erling Møller Pedersen.
Vejledere: Bernhelm Booss & Klaus Grünbaum
- 66/83 "MATEMATISKE MODELLER FOR PERIODISK SELEKTION I
ESCHERICHIA COLI".
Projektrapport af Hanne Lisbet Andersen, Ole
Richard Jensen og Klavs Frisdahl.
Vejledere: Jørgen Larsen og Anders Hede Madsen

67/83 "ELIPSOIDE METODEN - EN NY METODE TIL LINEÆR
PROGRAMMERING?"

Projektrapport af Lone Biilmann og Lars Boye

Vejleder: Mogens Brun Heefelt