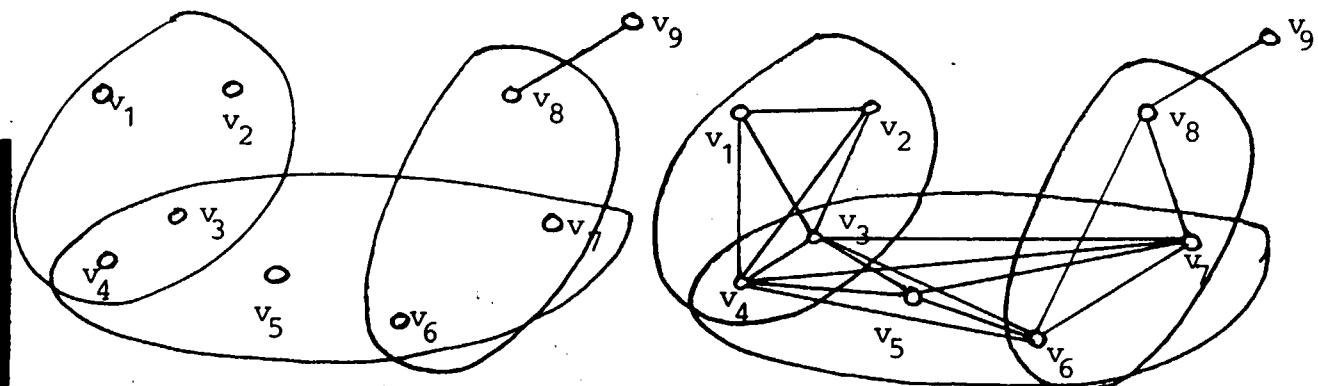


TEKST NR 114

1985



ANVENDELSE AF GRAFISKE METODER
TIL ANALYSE AF KONTINGENSTABELLER

vævtype \ placering	hoved, hals	krop	arme, ben	total
fregner	22	2	10	34
spredt overfladisk pigmentering	16	54	115	185
knuder	19	33	73	125
ubestemmelig	11	17	28	56
total	68	106	226	400

TEKSTER fra

IMFUFA

ROSKILDE UNIVERSITETSCENTER

INSTITUT FOR STUDIET AF MATEMATIK OG FYSIK SAMT DERES
FUNKTIONER I UNDERVISNING, FORSKNING OG ANVENDELSER

IMFUFA, Roskilde Universitetscenter, Postbox 260, 4000 Roskilde

ANVENDELSE AF GRAFISKE METODER TIL ANALYSE AF KONTINGENSTABELLER
Lone Biilmann, Ole R.Jensen og Anne-Lise von Moos.

IMFUFA tekst nr. 114/1985 60 sider ISSN 0106-6242

Abstract.

I rapporten undersøges, hvordan et matematisk fagområde -grafteori- kan anvendes i analyse af kontingenstabeller. Derfor gennemgås først den del af grafteorien herunder hypergraffer, som bruges i arbejdet med kontingenstabeller. Herefter beskrives den statistiske teori, der ligger til grund for analyse af kontingenstabeller.

Det viser sig, at grafteori kan bibringe analysen af kontingenstabeller en ny dimension, idet den grafteoretiske tilgang letter opstilling og udvælgelse af hypoteser. Den understøtter også den intuitive forståelse af sammenhænge/afhængighed, og gør det væsentligt lettere at bevare overblikket over de ellers ofte uoverskelige kontingenstabeller.

ANVENDELSE AF GRAFISKE METODER
TIL ANALYSE AF KONTINGENSTABELLER

LONE BIILMANN
OLE R. JENSEN
ANNE-LISE VON MOOS

VEJLEDER:
JØRGEN LARSEN

RUC 1985

INDHOLDSFORTEGNELSE.

0.	INDLEDNING.....	s.	1.
1.	MODELBEGBREBET.....	s.	2.
1.1.	Indledning.....	s.	2.
1.2.	Definition af model.....	s.	2.
1.3.	Forholdet mellem model og virkelighed.....	s.	3.
1.4.	Modeldannelsen.....	s.	3.
1.5.	Statistiske modeller for kontingenstabeller....	s.	7.
2.	GRAFTEORI.....	s.	9.
2.1.1.	Definition af en graf og dens diagram.....	s.	9.
2.1.2.	Rute, tur, vej længde og kreds.....	s.	12.
2.1.3.	Induceret del-graf, snitmængde og forbundethed.....	s.	14.
2.1.4.	Komplet graf, todelt graf og klike.....	s.	16.
2.1.5.	Rand og afslutning.....	s.	18.
2.2.	Hypergrafer.....	s.	19.
2.2.1.	Definition af hypergrafer.....	s.	19.
2.2.2.	Dekomposabel genererende klasse hypergraf....	s.	21.
2.2.3.	2-opdeling af hypergrafer og conformale hypergrafer.....	s.	24.
3.	KONTINGENSTABELLER.....	s.	26.
3.1.	Estimation generelt.....	s.	26.
3.1.1.	Model.....	s.	26.
3.1.2.	Maximum likelihood estimation.....	s.	27.
3.2.	Kontingenstabeller.....	s.	29.
3.2.1.	Kategorier og niveauer.....	s.	30.
3.2.2.	Indeksering af kontingenstabeller.....	s.	31.
3.2.3.	Analyse af en 2-dimensional kontingenstabel..	s.	32.
3.3.	Log-lineære modeller for kontingenstabeller..	s.	34.
3.3.1.	Log-lineære modeller for 2-dimensionale tabeller.....	s.	35.
3.3.2.	Første-ordens vekselvirkning.....	s.	37.
3.3.3.	Log-lineære tabeller for 3-dimensionale og multidimensionale tabeller.....	s.	38.
3.3.4.	Hierarkiske modeller.....	s.	39.
3.4.	Estimation i hierarkiske log-lineære modeller.....	s.	40.

3.4.1.	Valg af metode.....	s. 40.
3.4.2.	Principper for estimation.....	s. 41.
4.	GRAFER, HYPERGRAFER OG KONTINGENSTABELLER.....	s. 43.
4.1.	Indledning.....	s. 43.
4.1.1.	Mere om kontingenstabeller.....	s. 43.
4.1.2.	Markovfelter.....	s. 44.
4.1.3.	Sammenfatning af teorien for Markovfelter og kontingenstabeller.....	s. 46.
4.2.	Et eksempel på brug af grafer til test af en kontingenstabbel.....	s. 50.
5.	OPSAMLING OG AFRUNDING.....	s. 58.
	LITTERATURLISTE.....	s. 59.

0. INDLEDNING.

Dette er en projektrapport for 2 modul I-studerende (Ole R. Jensen + Lone Biilmann) og 1 modul II-studerende (Anne-Lise von Moos) ved matematikoverbygningsuddannelsen. Denne rapport skulle derfor dække både bredde, anvendelse samt modelkravet. Emnet "Grafteori's anvendelse ved analyse af kontingenstabeller" skulle opfylde dette krav, fordi vi behandler en teori (grafteori) som anvendes ved analyse af en klasse af statistiske modeller.

Da vi ved projektets start var interesserede i anvendelsen af grafteori, blev vi af vejleder Jørgen Larsen henvist til artiklen "Markov Fields and Log-Linear Interaction Models for Contingency Tables" (Darroch m.fl., 1980).

Artiklen dannede så grundlag for resten af projektforløbet. Denne rapport er så, hvad vi har nået at sæt^es os ind i om dette emne. Det er altså en sammenstykning af "gammelkendt" viden om grafteori og kontingenstabeller, samt den nyligt observerede anvendelse af grafteori som hjælpemiddel til analyse af kontingenstabeller. (Denne viden er så vidt vi ved ikke beskrevet i andet end artikelform).

Rapporten her omhandler de emner, der er nødvendige for at kunne forstå, hvordan en kontingenstabbel kan fortolkes v. hj. a. grafer, som for det første visualiserer en sådan model på en helt ny måde, for det andet indskrænker det antal af modeller, der analyseres, betragteligt, samtidig med at det kun er i sjældne tilfælde at interessante analyser overspringes.

1. MODELBEGREBET.

1.1. Indledning.

I dette afsnit gennemgåes nogle generelle overvejelser om matematiske modeller. Disse overvejelser går ud på at af-dække, hvad en matematisk model egentligt er. Endvidere beskrives i overordnede træk faserne i en modelbygning. Af-snittet afsluttes med en kort diskussion af kontingenstabellers anvendelse som grundlag for statistiske modeller.

1.2. Definition af model.

En model er en forenklet udgave af virkeligheden. Der kan være flere grunde til at lave en model. Den genstand eller det system, man modellerer, kan f.eks. være for kompliceret, farligt, langsomt, hurtigt, stort, lille eller udforsket til at eksperimentere med. Med en model kan man transformere sit genstandsområde til f.eks. et matematisk system, som der kan arbejdes med. Denne transformation består bl.a. i at udvælge de for emnet vigtige karakteristika.

For overskuelighedens skyld vil vi nu kalde de karakteristika, modellen består af, for elementer, og de relationer og sammenhænge, der er mellem disse, kalder vi for koblinger.

En model består således af en samling elementer samt nogle koblinger mellem disse elementer. Hvert af modellens elementer repræsenterer et element fra virkelighedsområdet, mens koblingerne ikke alle nødvendigvis repræsenterer kendte koblinger fra virkeligheden; der kan forekomme tænkte koblinger (Hermann m.fl., 1982).

Vi kan nu give en meget simpel definition på, hvad en mate-

matisk model er:

En matematisk model, er en model, hvor elementerne repræsenteres ved matematiske objekter, og koblingerne mellem elementerne oversættes til matematiske koblinger mellem de matematiske objekter.

1.3. Forholdet mellem model og virkelighed.

I enhver model er forholdet mellem model og virkelighed afgørende for troværdigheden af modellen. Det drejer sig ikke nødvendigvis kun om, hvor godt modellen stemmer overens med virkeligheden, mindst lige så vigtigt er det, om modellen er teoriunderbygget eller ej.

"Forhåndsforventningerne til modellens kvalitet må være større, hvis de forhold, der er taget i betragtning ved udvælgelsen og repræsentationen af objekter og strukturer, er i overensstemmelse med teorier eller teoridele, som fra andre sammenhænge er solidt underbygget, end hvis der er tale om til lejligheden konstruerede antagelser om virkelighedens beskaffenhed." (Birkemose, 1975)

For matematiske modeller gælder endvidere, at man i kraft af det matematiske sprog kan manipulere med modellens objekter og strukturer alene ud fra kendskabet til matematikkens symbolsprog, uden at ændre modellens udsagnskraft.

1.4. Modeldannelsen.

For at komme lidt mere i dybden med, hvad en matematisk model egentlig er, vil vi se på, hvordan man opstiller eller bygger en model.

Selve modelbygningen foregår som en proces og for at understrege dette, inddeler vi modelbygningen i nogle faser. Derned er det ikke sagt, at enhver modelbygning passerer de

enkelte faser i netop den rækkefølge, vi nævner dem, idet en sådan opstilling af modelbygningen i faser er en idealisert beskrivelse. Dette underbygger Hermann og Niss, som vi blandt andet støtter os til i dette afsnit, idet de om modelbygningens faser skriver:

"I virkeligheden foregår modelbygningen ofte langt mere uoverskueligt, med faserne vævet ind i hinanden, rækkefølgen mellem dem byttet om osv. Men det er alligevel sådan, at de ingredienser, der er beskrevet for den enkelte fase, alle er direkte eller inddirekte til stede i enhver model-dannelsessituation." (Hermann m.fl., 1982)

Resten af afsnittet omhandler de 4 faser, vi har inddelt modelbygningen i. De første 3 faser beskriver selve modelbygningen, mens den sidste fase omhandler behandlingen og evalueringen af modellen.

1. fase: udvælgelsen:

Som tidligere nævnt har modelbyggerens formål eller hensigt med at bygge en model en afgørende indflydelse på udvælgelsen af, hvilke elementer fra virkelighedsområdet, der i forbindelse med et opstillet problem er interessante at beskæftige sig med. Det er også modelbyggeren, der afgør hvilke koblinger mellem elementerne, der må betragtes som vigtige at medtage i modellen.

Denne fase består således af en udvælgelse af de elementer og koblinger, der skal indgå i modellen, og det fremgår klart, at denne fase for modelbyggeren indeholder mange valgsituationer, der ikke blot er afgørende for, hvordan den endelige model kommer til at se ud, men også er afgørende før modellens udsagnskraft.

Det er også i denne fase, det første informationstab forekommer; kun ved meget små og simple systemer kan man undgå et informationstab. Informationstabet behøver dog ikke at

være af særlig stor betydning, idet det kan hidrøre fra elementer, der ikke har betydning i denne sammenhæng.

2. fase: idealisering:

I denne fase foretages en idealisering af de i fase 1 udvalgte elementer og koblinger. Denne idealisering er ifølge Hermann og Niss, en

"...erstatning af den 'rå' virkelighed med en bearbejdet og mere poleret 'virkelighed'."

Omformningen af virkeligheden sker ved, at de udvalgte elementer og koblinger tilpasses på en sådan måde (dvs. omformes eller bortkastes), at en senere modeldannelse kan ske på en rimeligt overskuelig og overkommelig måde. Typiske idealiseringer er f.eks. gennemsnitsbetragtninger og små tilnærmelser og bestemmelse af, hvorvidt parametre er konstanter, eller om de snarere burde skrives som variable, der afhænger af systemets tilstand.

Ved omformningen eller bortkastelsen af elementer og koblinger sker et ægte informationstab, da disse elementer og koblinger rent faktisk har en betydning for den del af virkeligheden, modellen skal beskrive; men man beslutter på baggrund af en eller anden antagelse om, hvordan virkeligheden er indrettet, at nedprioritere visse faktorer for overhovedet at nå frem til en model, det er muligt at arbejde med.

Det skal dog bemærkes, at de informationstab der er sket indtil nu, er sket inden matematikken er inddraget i modellen.

3. fase: den matematiske repræsentation:

I denne fase drejer det sig om at oversætte de idealiserede elementer og koblinger til matematiske objekter og relationer. Hermann og Niss fremhæver i denne forbindelse, at

"Ved den matematiske repræsentation gøres en række matematiske forudsætninger om de oversatte objekter og koblingerne mellem dem."

De pointerer videre, at ikke alle disse forudsætninger svarer til de antagelser, der blev gjort om virkeligheden under udvælgelsen og idealiseringen, nogle antagelser gøres simpelthen for at den matematiske model overhovedet kan formuleres eller behandles.

Oversættelsen til det matematiske sprog kan ofte gøres på flere måder, så valget mellem disse forskellige måder må ske ud fra overvejelser om, hvad der gør det videre forløb bedst muligt. Når oversættelsen er tilendebragt, er modellen opstillet.

4. fase: behandlingen af modellen:

Det første, der sker ved behandlingen af modellen, er ofte at undersøge, om modellen overhovedet er "god nok", idet selve opstillingen af modellen kan resultere i noget uoverskueligt og ganske uhåndterligt rent matematisk. Det kan derfor være nødvendigt med nogle justeringer eller forandringer i den matematiske model. Dette kan gøres enten ved at foretage en simplificering af modellen eller ved at foretage nogle approximationer i den opstillede model.

Ved simplificeringen af modellen gøres

"...nogle supplerende matematiske forudsætninger, der indskrænker modellens generalitet til noget mere specifikt." (Hermann m.fl., 1982)

Denne simplificering medfører rent faktisk, at man erstatter den først opstillede model med et specialtilfælde af denne.

Ved approximation i modellen forstås døt, at tilnærme uhåndterlige matematiske udtryk og funktioner til mere håndterlige begreber. En sådan tilnærmlse kan dog give anledning til fejl i modellen, da det informationstab, der sker

ved en tilnærmelse til tider betragtes som tab i præcision og man taler om, at der opstår en usikkerhed. En tilnærmet model vil derfor ofte blive opfattet som mindre teoretisk tilfredsstillende end modellen uden tilnærmelser.

Når den matematiske behandling af modellen er tilendebragt, sammenlignes de fundne konklusioner med den bagvedliggende teori (hvis en sådan findes) eller med et givet sæt data.

Ved de teoretisk velfunderede modeller vil man ved uoverensstemmelse mellem teorien og resultatet blot forkaste resultatet med begrundelsen: der må være en fejl i udregningerne eller lignende. Denne type model evalueres derfor ikke som en enkeltstående model, men som led i en samlet teori, hvor det er teorien, man evaluerer snarere end selve modellen.

For de empirisk baserede modeller vil man derimod blot korrigere nogle af de i modellen indgående parametre, således at resultatet kommer til at stemme overens med allerede kendte data. Denne type model vil derfor blive evalueret som en enkeltstående model gennem konfrontation med et større eller mindre observationsmateriale.

Det er selvfølgelig de færreste modeller, der kan sættes i bås som enten 'teoretisk baseret' eller 'empirisk baseret' model. De allerfleste modeller er en blanding af disse to yderpunkter. Den skarpe opsplitning skal blot medvirke til at pointere, at

"Forholdet mellem matematiske modeller og den virkelighed, de repræsenterer, er ikke det samme for alle modeller." (Hermann m.fl., 1982)

Det er derfor vigtigt ved evaluering af enhver matematisk model at gøre sig dette forhold klart.

1.5. Statistiske modeller for kontingenstabeller.

De hypoteser, man opstiller for kontingenstabeller, kan opdeles i 2 grupper. Den ene gruppe består af rent statistiske

overvejelser, altså hypoteser om matematiske relationer mellem modellens variable. Den anden gruppe hypoteser hentes fra det fagområde de observerede data beskriver, og disse hypoteser prøver at give en forklaring på tingenes sammenhæng. Arbejdsgangen kan lidt firkantet beskrives som følger: Nogle fænomener er observeret, og det ville være interessant at be/afkræfte, om der er en sammenhæng mellem fænomenerne. Disse sammenhænge be- eller afkræftes statistisk, og man søger så i det relevante fagområde efter en forklaring på fortolkning af de påviste sammenhænge.

Analyse af kategoriske variable er blevet meget udbredt med billiggørelsen af datakraft. Det skyldes, at der er mange sammenhænge, der skal undersøges, hvilket kræver et stort antal beregninger. Men en analyse af kontingenstabeller er selvfølgelig ikke altid tilstrækkelig. Det er strukturen for svag til. Man sammenligner temmeligt ustruktureret et observeret kategorisk materiale for at finde en sammenhæng. En ukritisk brug af de fundne sammenhænge kan i visse tilfælde give resultater, der, når de 'konfronteres med virkeligheden', viser sig at være meningsløse. Man må altså her, som man iøvrigt altid bør gøre det, sammenholde modellens resultater med det virkelighedsområde, der skal modelieres.

2. GRAFTEORI.

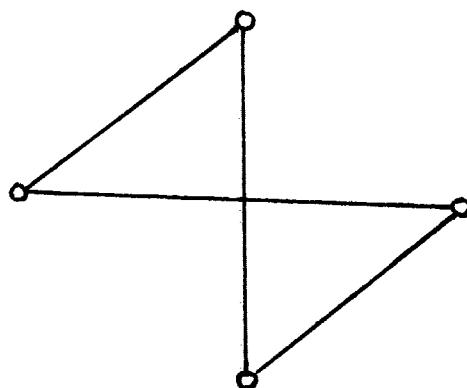
Afsnittet falder i to dele. Et om almindelige grafer og et om hypergrafer. For hele afsnittet gælder, at det er en beskrivende indføring i grafteori, dvs. beviser er udeladt.

Den første del, der omhandler almindelige grafer, beskæftiger sig kun med ikke-orienterede grafer. Den ikke-orienterede graf defineres, og forskellige grafteoretiske begreber gen nemgåes. Formålet med denne del af afsnittet er at introducere grafbegrebet bredt, samt at det skal virke som grundlag for de i anden del indførte hypergrafer.

Hypergrafer, som anden del beskæftiger sig med, omhandler definitioner samt begreber, der er nyttige for at forstå den grafiske behandling af kontingenstabeller.

2.1.1. Definition af en graf og dens diagram.

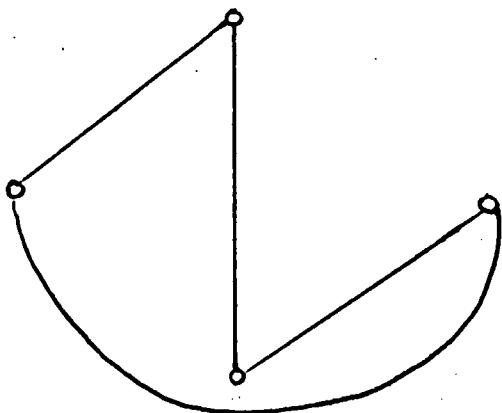
I mange situationer f.eks. indenfor trafik og mikro-teknologi kan det være hensigtsmæssigt at betragte problemerne som grafer. Herved man problemet som en række knuder (punkter), mellem hvilke der enten er forbindelse eller ikke-forbindelse. Dette kan illustreres vha. et diagram:



Figur 2.1.

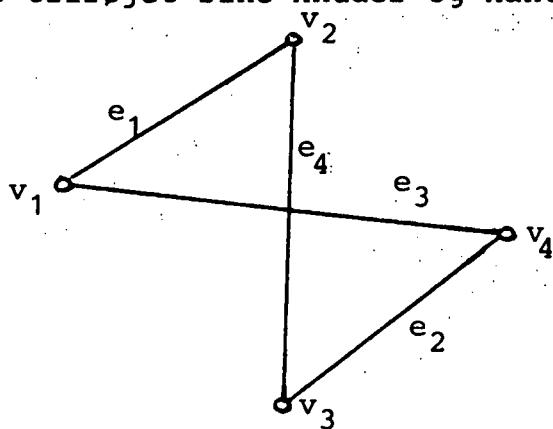
Det er den måde en graf i dette afsnit vil blive repræsenteret på. Denne graf består af 4 knuder og 4 kanter (forbindelser). Repræsentationen er ikke entydig, for diagrammet kunne også

have set således ud:



Figur 2.2.

En graf, for hvilken det er muligt at tegne et diagram, uden nogen af kanterne krydser hinanden, kaldes en plan graf. For at kunne referere til enkelte knuder/kanter/dele af grafen navngives disse. Grafen fra før kunne f.eks. have følgende navne tilføjet sine knuder og kanter:



Figur 2.3.

Figur 2.1, figur 2.2 og figur 2.3 er isomorfe, idet de alle repræsenterer den samme struktur. Knuderne og kanterne er de samme, blot er figur 2.3 navngivet. Figurerne 2.1 og 2.2 siges at repræsentere en ækvivilens-klasse af isomorfe grafer, der kunne være navngivet på en hvilken som helst måde.

Nu kan der opstilles en definition af grafen G i figur 2.3:

$$G = (V(G), E(G), \gamma_G)$$

hvor

$$V(G) = \{v_1, v_2, v_3, v_4\}$$

$$E(G) = \{e_1, e_2, e_3, e_4\}$$

og ψ_G er defineret ved

$$\psi_G(e_1) = v_1v_2, \psi_G(e_2) = v_3v_4$$

$$\psi_G(e_3) = v_1v_4, \psi_G(e_4) = v_2v_3.$$

Man kan sige at grafen G er defineret ved en tripel, nemlig mængderne $V(G)$ og $E(G)$ og afbildningen ψ_G .

Forklaring af triplen:

$V(G)$: er mængden af knuder i grafen G
- fremover kaldet V.

$E(G)$: er mængden af kanter i grafen G
- fremover kaldet E.

ψ_G : er en funktion, som for hver kant angiver kantens endeknuder.

Ofte skelnes der ikke formelt mellem ψ_G og $E(G)$, idet f.eks. e_1 kan betegnes ved:

$$e_1 = \{v_1, v_2\}.$$

Generelt kan en kant på denne måde defineres ved:

$$e_k = \{v_i, v_j\};$$

Denne repræsentation kan ikke angive dobbeltkanter!

Endvidere bruges symbolerne $|V|$ og $|E|$ til at betegne antallet af knuder hhv. kanter i grafen G. I vores eksempel er

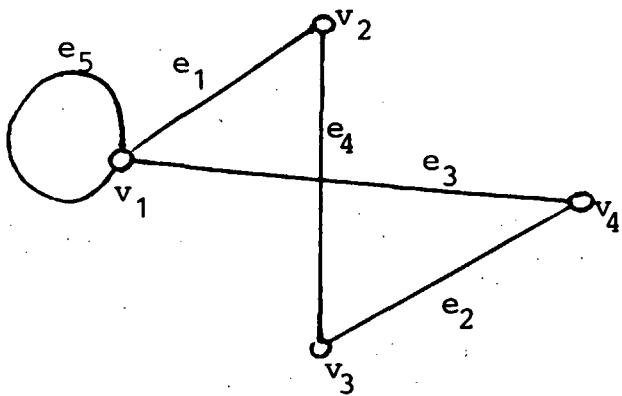
$$|V| = |E| = 4$$

og to knuder siges at være naboer, hvis der er en kant mellem dem.

Hvis der tilføjes en ny kant e_5 , hvor

$$e_5 = \{v_1, v_1\}$$

ville diagrammet se således ud:



Figur 2.4.

En kant som e_5 kaldes en løkke. En løkke er altså en kant, hvis afbildning kun indeholder en enkelt knude - eller sagt på en anden måde:

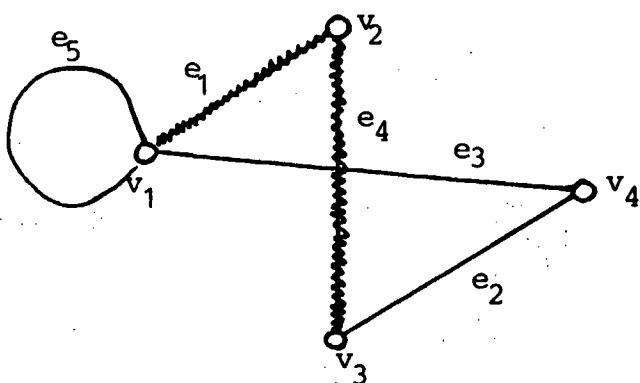
$$\text{en kant } e_k = \{v_i, v_i\} .$$

En graf er simpel, hvis den ikke indeholder løkker og to givne knuder ikke er forbundet af mere end en kant. Figur 2.5 viser et eksempel på en ikke-simpel graf.

2.1.2. Rute, tur, vej længde og kreds.

Hvis grafen fra før igen betragtes kan man f.eks. lave følgende rute R :

$$R = v_1 e_1 v_2 e_4 v_3$$



Figur 2.5.

En rute er altså et gennemløb i hele eller dele af grafen.

Mere abstrakt kan en rute opskrives på følgende måde:

$$R=v_0e_1v_1\ldots\ldots e_kv_k$$

For vores konkrete rute fra før kan vi "oversætte" den formelle notation som eksempel:

Konkret graf:

v_1
 e_1
 v_2
 e_4
 v_3

formel notation:

v_0
 e_1
 v_1
 e_2
 v_2

En rute kan kaldes en tur, hvis én betingelse er opfyldt. Nemlig den, at alle kanterne i ruten skal være forskellige, mens knuderne godt må passeres flere gange. Man kan også tildele en tur en længde. En tur's længde er defineret som antallet af kanter, turen indeholder.

En mere restriktiv måde at 'vandre' rundt i en graf på kaldes en vej. En vej er en tur, hvor også alle knuderne er forskellige. Hvis grafen fra før igen betragtes, er der herunder angivet et eksempel på de tre måder at bevæge sig rundt i en graf på. Først den mindst restriktive, nemlig ruten R så en tur T og til slut en vej P.

$$R=v_1e_5v_1e_1v_2e_4v_3e_2v_4e_3v_1e_5v_1$$

$$T=v_4e_3v_1e_5v_1e_1v_2$$

$$P=v_3e_3v_4e_2e_1v_1e_3v_4$$

Den formelle opskrivning for en rute betragtes igen:

$$V=v_0e_1v_1e_2\ldots\ldots e_kv_k$$

v_0 kaldes begyndelsesknode og v_k terminalknoden. Alle de mellemliggende knuder kaldes interne knuder.

For en tur gælder der, at hvis begyndelsesknode og terminalknuden er den samme, hedder turen en kreds. En kreds kan selvfølgelig også tildeles en længde og en kreds af længden k kaldes en k -kreds. Specielt kaldes en 3-kreds for en tre-

kant. Fra den snart meget brugte graf fra tidligere er her et eksempel på en 4-kreds:

$$K=v_1e_3v_4e_2v_3e_4v_2e_1v_1$$

Både ruter, ture og veje kan danne en kreds.

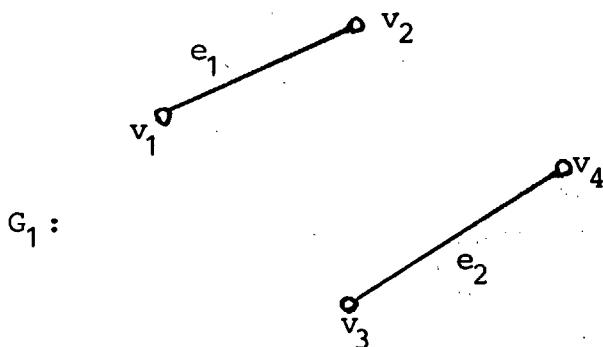
2.1.3. Induceret del-graf, snitmængde og forbundethed.

Mange begreber fra mængdelæren anvendes også indenfor grafteorien. Et eksempel på dette er begrebet en del-graf. En del-graf til en graf G defineres i forhold til G og delgrafen G_1 skal opfylde følgende i forhold til G :

$$V(G_1) \subseteq V(G)$$

$$E(G_1) \subseteq E(G)$$

Herunder gives et eksempel på en delgraf G_1 af den gammelkendte graf G .

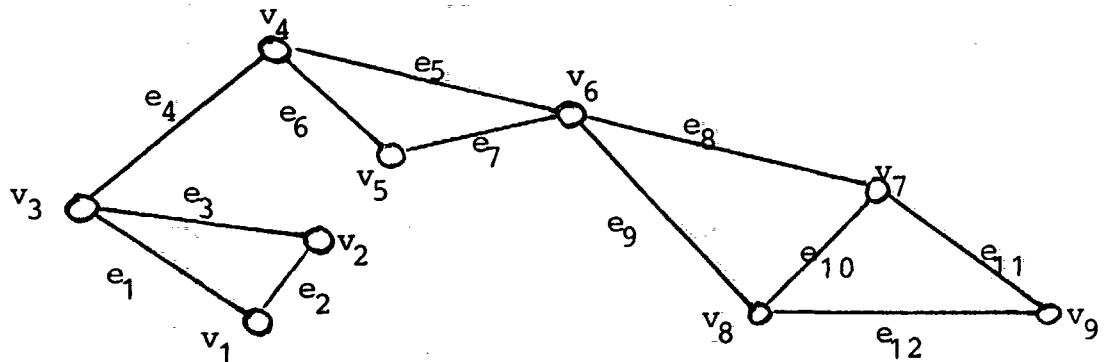


Figur 2.6.

I delgrafen G_1 er v_1 og v_2 forbundet af kanten e_1 og v_3 og v_4 af e_2 .

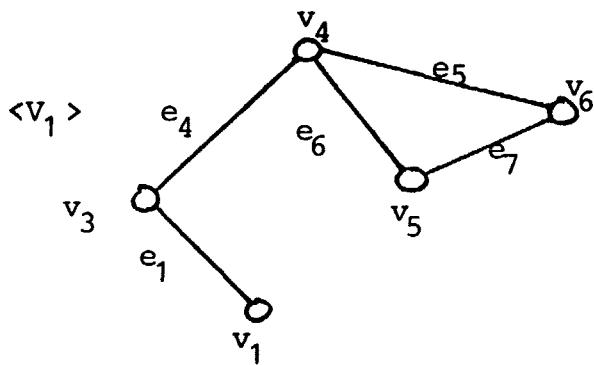
Grafen i figur 2.6 siges at være ikke-forbundet i modsætning til G , der er forbundet. En forbundet graf er en graf, hvor der findes en vej mellem to vilkårlige knuder. Hvis en graf ikke er forbundet kaldes de enkelte dele for grafens sammenhængskomponenter, dvs. en hver delgraf bestående af én sammenhængskomponent er forbundet.

En snitmængde er en mængde af knuder (eller kanter) hvis fjernelse resulterer i en ikke forbundet graf, dvs. f.eks. er knudemængden $\{v_4, v_5\}$ i figur 2.7 en snitmængde.



Figur 2.7.

Man kan lave en speciel slags delgraf, kaldet reduceret delgraf. Vi tager grafen i figur 2.7 og betragter en delmængde V_1 af V . Vi sætter $V_1 = \{v_1, v_3, v_4, v_5, v_6\}$. Så findes den(de) kant(er), der har begge endeknuder i V_1 - her kanterne e_1, e_4, e_5, e_6 og e_7 . Den reducerede delgraf betegnes med $\langle V_1 \rangle$ som er vist på figur 2.8.

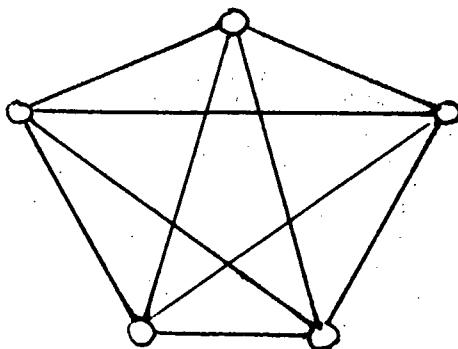


Figur 2.8.

En snitmængde deler en graf i nogle sammenhængskomponenter. Delgrafen reduceret ved knuderne i en sammenhængskomponent forenet med knuderne i snitmængden kaldes et stykke af grafen. Stykkerne af en graf er samtlige delgrafer, der fremkommer ved denne procedure.

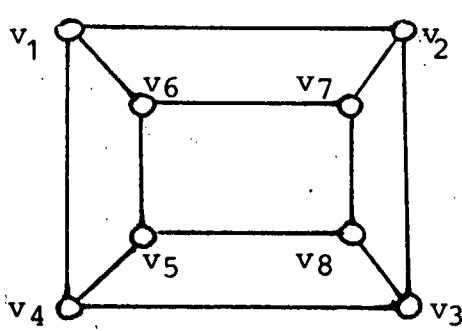
2.1.4. Komplet graf, todelt graf og klike.

En graf, hvor alle forskellige par af knuder er forbundet med en kant, kaldes en komplet graf. Der findes kun én komplet graf for hvert n (alle komplette n -grafer er isomorfe). Denne betegnes K_n , hvor n er antallet af knuder. F.eks. ser K_5 således ud:

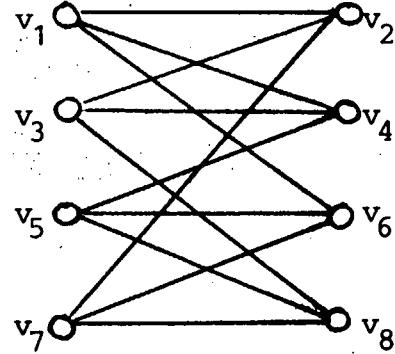


Figur 2.9.

En todelt graf er en graf, hvor knudemængden kan deles i to delmængder. De to delmængder består af knuder, sådan at alle kanterne har en endeknude i hver delmængde. Et eksempel på en todelt graf følger herunder:



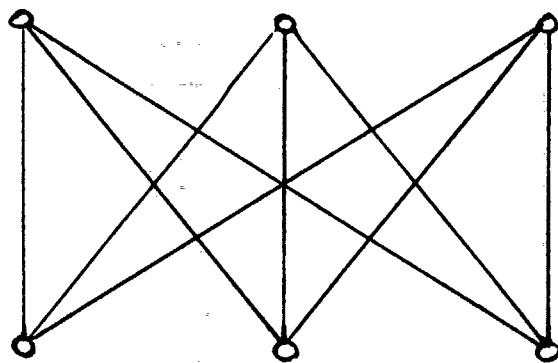
Figur 2.10 (a).



Figur 2.10 (b).

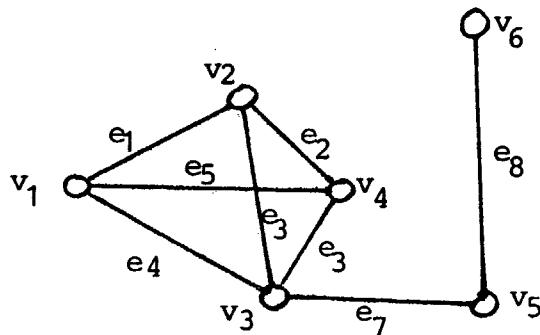
De to grafer er isomorfe. Det er tydeligt i figur 2.10(b), at grafen er todelt.

De to begreber todelt og komplet kan samles i et krav til en graf, nemlig den todelte komplette graf. Her følger et eksempel på en sådan:



Figur 2.11.

Det ses af figuren, at denne type graf ikke er komplet - men at den derimod er komplet mht. sin todeling. Dvs. at hver knude i den delmængde, der fremkommer ved todelingen, er 'nabo' til alle knuderne i den anden delmængde.



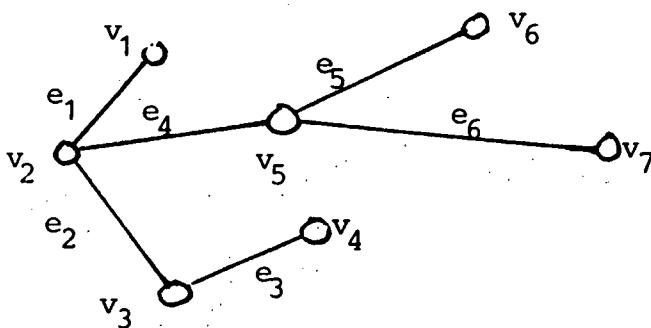
Figur 2.12.

Hvis grafen i figur 2.12 betragtes, ses det, at den ved delmængden $V_1 = \{v_1, v_2, v_3\}$, inducerede delgraf $\langle V_1 \rangle$ er en komplet graf. Det samme gælder for $\langle V_2 \rangle$, hvor $V_2 = \{v_1, v_2, v_3, v_4\}$. Den sidste kaldes en klike, fordi den er maksimal. Dvs., at ligegyldigt, hvilken anden knude, der tilføjes delgrafen, vil den så fremkomne delgraf ikke mere være komplet.

2.1.5. Rand og afslutning.

Hvis man ser på følgende graf G med delgrafen $\langle v_1 \rangle$ med

$$v_1 = \{v_1, v_2, v_5\}$$

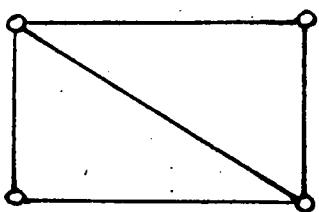


Figur 2.13.

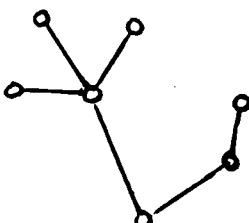
kaldes knuderne v_3, v_6 og v_7 for randen til G_1 eller ∂G_1 . Randen (∂G_1) til en graf G er altså defineret som naboknuderne til delgrafen. Hvis man har en delgraf G_1 , så er afslutningen til G_1 foreningsmængden af G_1 og dens rand.

2.1.6. Triangulerede grafer.

En graf er triangulær, hvis den ikke indeholder en kreds med længde større end eller lig 4 uden en korde (en kant der ikke ligger på randen, men som forbinder to knuder). Figur 2.14 viser to triangulerede grafer.

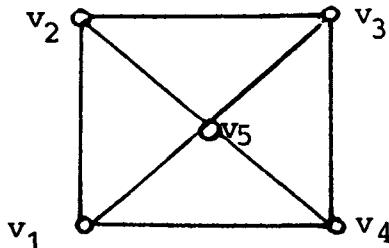


Figur 2.14(a).

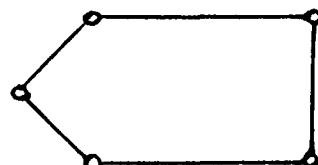


Figur 2.14(b).

Figur 2.14 (a) indeholder en kreds med længde 4, men den har en korde. Grafen i figur 2.14(b) er trianguleret, da den ikke indeholder en kreds (en sådan graf kaldes et træ).



Figur 2.15(a).



Figur 2.15(b).

Derimod er graferne i figur 2.15 ikke trianguleret. 4-kredsen v_1, v_2, v_3 og v_4 har nemlig ingen korde. Figur 2.15(b) indeholder kun en kreds, og da længden er 5, er grafen ikke trianguleret.

2.2. Hypergrafer.

For senere at kunne tydeliggøre sammenhængen mellem grafer og statistiske modeller, indføres nu hypergrafer, der kan betragtes som en udvidelse af grafbegrebet. I en almindelig graf kan man definere kanterne som en mængde af uordnede par af knuder. Som det fremgår af nedenstående, består udvidelsen af grafbegrebet til hypergrafer i, at knudemængden, der definerer en kant i en hypergraf, kan indeholde et vilkårligt antal knuder (>0).

2.2.1. Definition af hypergrafer.

En hypergraf defineres (Berge, 1973) ud fra en mængde af knuder

$$V = \{v_1, v_2, \dots, v_n\}$$

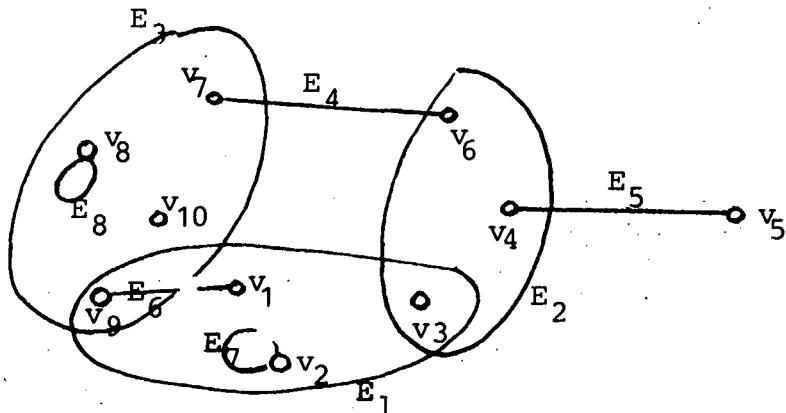
og en klasse af delmængder af V : $\mathcal{E} = \{E_i \mid i \in I\}$, hvor I er

en indexmængde, således at

$$E_i \neq \emptyset \quad (i \in I)$$

$$\bigcup_{i \in I} E_i = V.$$

Hypergrafen betegnes $H = (V, \mathcal{E})$, og mængderne E_1, E_2, \dots, E_m kaldes hypergrafens kanter.



Figur 2.16.

Hypergrafens orden betegnes $|V|$ og angiver antallet af knuder. Her er altså $|V| = 10$. Tilsvarende betegnes antallet af kanter $|\mathcal{E}| (= |I| = m)$ og antallet af knuder i en kant med $|E_i|$. Af figur 2.16 ses, at hvis $|E_i| = 1$ symboliseres kanten med en løkke, er $|E_i| = 2$ ved en linie, og er $|E_i| > 2$, tegnes en kurve omkring E_i 's knuder.

For hypergrafer kan man benytte begreber/betegnelser, der er analoge til dem, vi benytter i "almindelige" grafer (fremover blot kaldet grafer), hvad vi viser eksempler på i det følgende.

En simpel hypergraf er en hypergraf, hvor $E_i \neq E_j$ hvis $i \neq j$; to knuder er naboyer, hvis de tilhører samme kant, og to kanter er naboyer, hvis deres fællesmængde ikke er tom. Det fremgår, at hvis $|E_i| = 2$ for alle i , og hypergrafen er simpel, da vil den være isomorf med en simpel graf uden isolerede knuder. Sammenhængskomponenter, vej og kreds er også analoge for grafer og hypergrafer: Man opskriver en vej på følgende måde:

$v_0 E_1 v_1 E_2 \dots E_p v_p$, hvor $v_i \neq v_j$, $E_i \neq E_j$ for $i \neq j$

og $v_{k-1}, v_k \in E_k$.

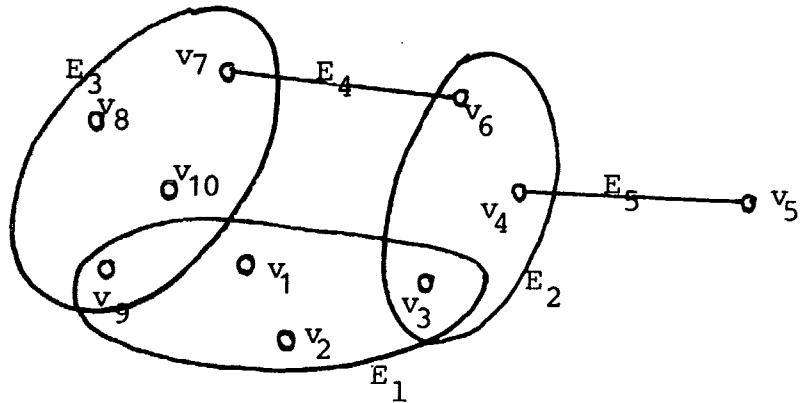
Erl $v_0 = v_k$ er der tale om en kreds. Et eksempel herpå fra hypergrafen i figur 2.16 er

$$v_1 E_6 v_9 E_3 v_7 E_4 v_6 E_2 v_3 E_1 v_1.$$

Endvidere er begreber som snitmængde, induceret delhypergraf og stykker af hypergrafen analoge til de tilsvarende betegnelser for grafer.

2.2.2. Dekomposabel genererende klasse hypergraf.

Idet følgende introduceres nogle nye begreber, og der opskrives nogle sætninger og definitioner for genererende klasse hypergrafer (g.k. hypergrafer): En g.k. hypergraf er en simpel hypergraf, der yderligere opfylder betingelsen $E_i \not\subseteq E_j$ for $i \neq j$. Figur 2.17 viser en g.k. hypergraf, der er en delhypergraf af den i figur 2.16 viste hypergraf.



Figur 2.17.

De nedenfor anførte nummererede sætninger og definitioner findes (ofte i en mere generel form) hos Lauritzen m.fl. (1978) som definitioner (d), corolarer (c), lemmaer (l) eller teoremer (t). Vi skal imidlertid først have indført en særlig type fællesmængde, betegnet Λ , for stykker af en g.k. hypergraf (V, \mathcal{E}) , der er delt i to stykker (V_1, \mathcal{E}_1) og (V_2, \mathcal{E}_2) , hvor $V_1, V_2 \subseteq V$ og $\mathcal{E}_1, \mathcal{E}_2 \subseteq \mathcal{E}$. Desuden betegner vi fremover hypergrafen ved angivelse af mængden af kanter $\mathcal{E} (\mathcal{E}_1, \mathcal{E}_2 \text{ osv.})$. Vi indfører så, at fællesmængden beteg-

net ved $\xi_1 \wedge \xi_2$ er klassen af maksimale elementer bestående af $E_i \cap E_j$ for alle $E_i \in \xi_1$ og $E_j \in \xi_2$.

Vi kan så definere en dekomposition (her: i to stykker) af en g.k.hypergraf:

1 (d) En g.k.hypergraf bliver dekomponseret i ξ_1 og ξ_2 i forhold til (en snitmængde) $d \subseteq V$, hvis $\xi_1 \cup \xi_2 = \xi$ og $\xi_1 \wedge \xi_2 = \{d\}$.

For at anskueliggøre dette ser vi på g.k.hypergrafen i figur 2.17. Den kan dekomponseres i forhold til $d = \{v_4\}$, idet vi så får $\xi_1 = \{E_1, E_2, E_3, E_4\}$ og $\xi_2 = \{E_5\}$. Det ses, at $\xi_1 \cup \xi_2 = \xi$ og hvis vi indledningsvis ser bort fra kravet om maksimalitet, fås

$$\xi_1 \wedge \xi_2 = \{E_1 \cap E_5, E_2 \cap E_5, E_3 \cap E_5, E_4 \cap E_5\} = \{\{4\}, \emptyset\}$$

idet $E_2 \cap E_5 = \{v_4\}$ og de øvrige fællesmængder er tomme. Medtager vi kravet om maksimalitet, fås $\xi_1 \wedge \xi_2 = \{\{4\}\} = \{d\}$. ξ kan altså dekomponseres i forhold til $d = \{4\}$.

ξ kan derimod ikke dekomponseres i forhold til $d = \{v_3, v_7\}$. Stykkerne af ξ bliver $\xi_1 = \{E_1, E_3\}$ og $\xi_2 = \{E_2, E_4, E_5\}$. Ser vi igen bort fra maksimalitet, fås

$$\xi_1 \wedge \xi_2 = \{E_1 \cap E_2, E_3 \cap E_2, E_1 \cap E_4, \dots, E_3 \cap E_5\} = \{\{v_3\}, \{v_7\}, \emptyset\}.$$

Indrages maksimalitet, fås

$$\xi_1 \wedge \xi_2 = \{\{v_3\}, \{v_7\}\} \neq \{\{v_3, v_7\}\}$$

g.k.hypergrafen i figur 2.17 kan altså ikke dekomponseres i forhold til $d = \{v_3, v_7\}$.

Vi kan så give en definition på en dekomposabel g.k.hypergraf:

2 (d) En g.k.hypergraf ξ er dekomposabel, hvis enten $|\xi| = 1$, eller hvis der findes en dekomposition af ξ i forhold til $d \subseteq V$ i ξ_1 og ξ_2 , så $|\xi_1| < |\xi|$, $|\xi_2| < |\xi|$ og både ξ_1 og ξ_2 er dekomposable.

Det ses, at den i figur 2.19 viste g.k.hypergrafen ikke er dekomposabel, idet ξ_1 , fremkommet i førstnævnte eksempel ikke er dekomposabel.

Vi kan så indføre begrebet ekstremal kant. Det skal nævnes, at dette begreb har størst betydning, hvis man er interesseret i at lave en algoritme til undersøgelse af, om en given g.k.hypergraf er dekomposabel.

3 (d) En ekstremal kant E^* i en g.k.hypergraf Σ er en kant, der opfylder betingelsen: Der findes en kant $E^{**} \in \Sigma \setminus \{E^*\}$ således at $E \cap E^* \subseteq E^{**} \cap E^*$ for enhver kant $E \in \Sigma \setminus \{E^*\}$.

Ser vi endnu en gang på figur 2.17, ses det, at E_5 er ekstremal ($E^* = E_5$), idet der findes en kant $E^{**} (= E_2)$, således at 3 er opfyldt: $E \cap E_5 = \emptyset$ hvis E er alle andre kanter end E_2 , og $E_2 \cap E_5 = \{v_4\}$, dvs. $E \cap E_5 = \emptyset \subseteq E_2 \cap E_5 = \{v_4\}$ for $E \neq E_2, E_5$ (betingelsen $E \cap E^* \subseteq E^{**} \cap E^*$ er naturligvis opfyldt, når man vælger $E = E^{**}$). En nærmere undersøgelse vil vise, at ingen andre kanter i denne hypergraf er ekstremale, f.eks. er E_4 ikke ekstremal: Vi sætter $E^* = E_4$ og vælger f.eks. $E^{**} = E_2$. Det giver $E^* \cap E^{**} = \{v_6\}$. Når vi så undersøger om $E \cap E^* \subseteq E^* \cap E^{**}$ fås for $E = E_3$: $E_3 \cap E^* = \{v_7\} \not\subseteq \{v_6\}$. Et tilsvarende resultat fås, hvis en anden kant vælges som E^{**} .

Der er to sætninger om ekstremale kanter i g.k.hypergrafer, der har betydning, hvis man som ovenfor nævnt er interesseret i en algoritme til at undersøge om en g.k.hypergraf er dekomposabel:

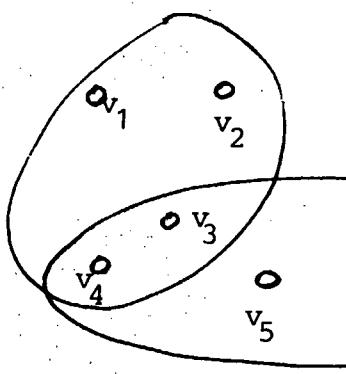
4 (c) Hvis E^* er en ekstremal kant i en dekomposabel g.k.hypergraf Σ , så er $\Sigma \setminus \{E^*\}$ dekomposabel, og

5 (1) En dekomposabel g.k.hypergraf Σ har mindst to ekstremale kanter (såfremt $|\Sigma| \geq 2$).

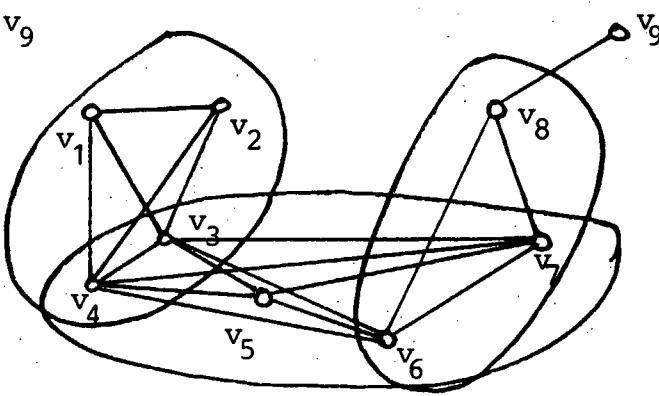
2.2.3. 2-opdeling af hypergrafer og conformale hypergrafer.

For at få undersøgelser af g.k. hypergrafer bragt over til undersøgelse af (alm.) grafer, kan man betragte 2-opdelingen ("2-section") af hypergrafer. 2-opdelingen af en g.k. hypergraf (V, ξ) kan defineres som grafen, der har V som knuder, og som kanter har mængden af (ikke-ordnede) knudepar $\{v_i, v_j\}$, hvor v_i og v_j begge tilhører samme kant i hypergrafen.

Figur 2.18 viser et eksempel på en hypergraf (figur 2.18(a)) og dens 2-opdeling (figur 2.18(b)).

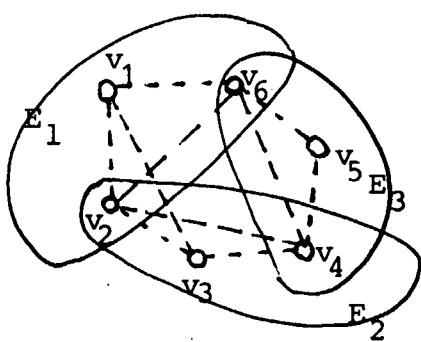


Figur 2.18(a).



Figur 2.18(b).

En conformal g.k. hypergraf (V, ξ) kan defineres som en g.k. hypergraf, hvor klassen af alle kliker af hypergrafens 2-opdeling netop udgør ξ . Hypergrafen i figur 2.17 og figur 2.18 er begge conformale. Et eksempel på en g.k. hypergraf, der ikke er conformal, er hypergrafen med $\xi = \{E_1, E_2, E_3\}$, hvor $E_1 = \{v_1, v_2, v_6\}$, $E_2 = \{v_2, v_3, v_4\}$ og $E_3 = \{v_4, v_5, v_6\}$, se figur 2.19. 2-opdelingen er antydet med punkterede linier.



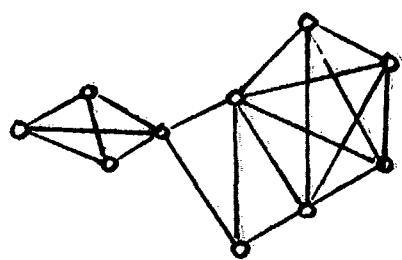
Figur 2.19.

Delgraferne reduceret ved knuderne i hhv. E_1, E_2 og E_3 er selvfølgelig klier, men delgrafen reduceret ved v_1, v_4 og v_6 er også en klike, men $\{v_2, v_4, v_6\} \notin \Sigma$, og Σ er derfor ikke conformal.

Det kan vises, at

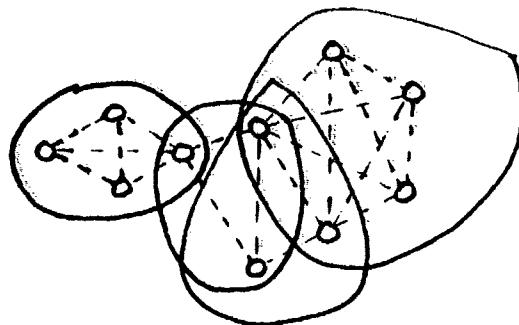
6 (c) Enhver dekomposabel g.k. hypergraf er conformal. Det omvendte er ikke nødvendigvis tilfældet, jvf. figur 2.17.

Alt dette giver anledning til at sammenligne grafer med klike-hypergrafer, hvor klikehypergrafen dannes af grafen ved at hypergrafens kanter består af knuderne i grafens klier, se figur 2.20. I klikehypergrafen (figur 2.20(b)) er grafens (figur 2.20(a)) kanter markeret ved punkterede linier.



En graf

Figur 2.20(a).



Dens klikehypergraf

Figur 2.20(b).

Det kan så vises, at for en sammenhængende graf er følgende punkter a og b ækvivalente:

7 (t) a Der findes ingen delmængde af V , der reducerer en delgraf med en kreds af længde > 3 , uden en korde.

b Klikehypergrafen er dekomposabel.

Betydningen af teoremet 7 vil fremgå i kapitel 4, hvor vi sammenkobler dele af teorien for grafer og hypergrafer med begreber fra statistikken.

3. KONTINGENSTABELLER.

Vi vil i dette kapitel beskrive, hvad en kontingenstabell er, og hvordan man analyserer den. Inden vi når frem til selve kontingenstabellerne, vil vi dog først komme lidt ind på estimation generelt samt mere specielt se på maximum likelihood estimation. Derefter vil vi ved hjælp af et eksempel beskrive de 2-dimensionale kontingenstabeller. Vi indfører begrebet log-lineære modeller for 2-dimensionale kontingenstabeller og udvider det til også at gælde for flere-dimensionale kontingenstabeller, og til sidst kommer vi ind på estimation i de log-lineære modeller.

3.1. Estimation generelt.

I det følgende afsnit kommer vi først ind på det mere generelle omkring estimering i statistiske modeller. Vi nævner to estimations metoder, og heraf uddyber vi den ene - maximum likelihood metoden.

3.1.1. Model.

Til analyse af en større mængde af data (observationer) anvendes en statistisk model knyttet til en sandsynlighedsmodel for at få en overskuelig beskrivelse af disse data og finde eventuelle sammenhænge/regelmæssigheder i den pågældende observation (datamængde). En sådan observation kan være fremkommet på flere forskellige måder, ofte i form af en stikprøve, dvs. udvælgelse af en delmængde af en population, et vareparti el.lign. (udvælgelsesstrategier vil vi ikke komme nærmere ind på).

De indsamlede data kan repræsenteres ved observationen

$$\underline{y} = (y_1, y_2, \dots, y_n),$$

og vi betragter \underline{y} som en realiseret observation af den sto-

kastiske variable

$$\underline{Y} = (Y_1, Y_2, \dots, Y_n),$$

og det drejer sig om at teste/fitte en sandsynlighedsmodel for den stokastiske variabel \underline{Y} overfor den givne observation. Vi betragter sandsynlighedsmodellen som specifiseret ved sandsynlighedsfunktionen $p(\underline{y}, \underline{\theta})$, og vi opfatter altså \underline{y} som en observeret værdi af den stokastiske variabel \underline{Y} , der følger sandsynlighedsfordelingen, som er bestemt af $p(\underline{y}, \underline{\theta})$, hvor $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ har en bestemt, men ukendt værdi.

$\underline{\theta}$ skal så estimeres i overensstemmelse med den foreliggende hypotese om sandsynlighedsfordeling og $\underline{\theta}$'s "udseende". Det skal ske således, at estimatoren, der betegnes $\hat{\underline{\theta}}$, bliver "den bedst mulige", der er i overensstemmelse med hypotesen og dens foreliggende observation \underline{y} . Der findes forskellige metoder til bestemmelse af $\hat{\underline{\theta}}$. Man kan f.eks. anvende maximum likelihood metoden, der omtales nærmere nedenfor, eller man kan for forskellige "bud" på $\underline{\theta}$ finde de $\hat{\underline{\theta}}$, der opfylder $E\hat{\underline{\theta}} = \underline{\theta}$, og blandt disse vælges så den med den mindste varians. Herved fås en middelret ($E\hat{\underline{\theta}} = \underline{\theta}$) og variansminimal estimator. De nærmere enkelheder ved denne estimationsmetode vil vi ikke komme ind på.

3.1.2. Maximum likelihood estimation.

Vi betragter som før nævnt \underline{y} som en observation f.eks. en stikprøve frembragt af sandsynlighedsfunktionen $p(\underline{y}, \underline{\theta})$. Betragter man en anden observation (stikprøve) af den samme $p(\underline{y}, \underline{\theta})$, vil man givetvis få en anden værdi for \underline{y} , og dermed et andet skøn over estimatoren $\hat{\underline{\theta}}$, dvs. $\hat{\underline{\theta}} = \hat{\underline{\theta}}(\underline{y})$, og maximum likelihood metoden er en metode til at finde det bedste skøn over $\hat{\underline{\theta}}$ ud fra den givne observation.

Vi indfører en funktion, likelihoodfunktionen:

$$L : \underline{\theta} \rightarrow p(\underline{y}, \underline{\theta})$$

$L=L(\underline{\theta}, \underline{y})$ angiver for enhver værdi af $\underline{\theta}$ "rimeligheden" af,

at $\underline{\theta}$ har frembragt den givne observation \underline{y} (under den opstillede model), med andre ord: jo større værdien af L er, desto bedre er skønnet over $\underline{\theta}$, og den kan bestemmes som:

$$L(\hat{\underline{\theta}}|\underline{y}) = \max_{\underline{\theta}} L(\underline{\theta}|\underline{y})$$

Man kan således finde maximum for $\underline{\theta}$ ved at differentiere L og undersøge før maximum på sædvanlig vis. Det vil dog ofte være nemmere at differentiere $\ln L(\underline{\theta}|\underline{y})$ (log er den naturlige logaritme, og vil være det fremover). Idet

$$\ln L(\underline{\theta}|\underline{y}) = \ln L(\underline{\theta}|\underline{y})$$

får man, hvis der er k parametre, der skal estimeres, k likelihoodligninger af typen

$$\frac{\partial \ln L(\underline{\theta}_i|\underline{y})}{\partial \underline{\theta}_i} = 0$$

Man må naturligvis sikre sig, at de fundne $\underline{\theta}_i$ er globale maxima. Metoden kan illustreres ved at tage et Bernoulli-forsøg som eksempel: $\underline{y} = (y_1, y_2, \dots, y_n)$, $y_i \in \{0, 1\}$. Punktsandsynligheden er

$$P(Y_i=y_i) = \underline{\theta}^{y_i} (1-\underline{\theta})^{1-y_i}$$

og sandsynlighedsfunktionen (den simultane punktsandsynlighed) bliver

$$p(\underline{y}|\underline{\theta}) = \prod_{i=1}^n P(Y_i=y_i) = \underline{\theta}^{\sum y_i} (1-\underline{\theta})^{n-\sum y_i}$$

Likelihoodfunktionen bliver

$$L(\underline{y}|\underline{\theta}) = \underline{\theta}^{\sum y_i} (1-\underline{\theta})^{n-\sum y_i}, \quad \underline{\theta} \in [0; 1]$$

og likelihoodligningen bliver

$$\frac{d(\sum y_i \ln \underline{\theta} - (n-\sum y_i) \ln(1-\underline{\theta}))}{d\underline{\theta}} = 0$$

$$\frac{\sum y_i}{\underline{\theta}} - \frac{n-\sum y_i}{1-\underline{\theta}} = 0$$

dvs. at vi for maximum likelihood estimatet (MLE) får

$$\hat{\underline{\theta}} = \frac{\sum y_i}{n}$$

Sandsynlighedsfunktionen kan ofte omskrives til

$$p(\underline{y}, \underline{\theta}) = g(s(\underline{y}), \underline{\theta}) \cdot h(\underline{y}) ,$$

hvor $h(\underline{y})$ er en af $\underline{\theta}$ uafhængig funktion, medens al information om $\underline{\theta}$ er indeholdt i $s(\underline{y})$, den såkaldte stikprøvefunktion.

Denne omskrivning kaldes en sufficient reduktion. Det ses i dette tilfælde, at $s(\underline{y}) = \sum y_i$. Det samme vil være tilfældet, hvis \underline{y} stammer fra en binomialfordelt stokastisk variabel Y , idet binomialfordelingen kun adskiller sig fra ovenstående fordeling ved binomialkoefficienten, der er indeholdt i $h(\underline{y})$.

Går vi over til at betragte en multinomialfordeling, vil det ssmme gøre sig gældende for hver enkelt θ_i i $\underline{\theta}$. Man kan endvidere vise, at noget tilsvarende gælder for en Poissonfordelt stokastisk variabel \underline{Y} , idet man også her får MLE (for λ) til

$$\hat{\lambda} = \frac{\sum y_i}{n}.$$

I tilfælde, hvor der anvendes log-lineære modeller, kan dette siges at være praktisk, idet man kan foretage estimering i kontingenstabeller, uden at man i selve estimationsprocessen behøver at tage hensyn til, om man i modellen har valgt en multinomial- eller en Poissonfordeling.

3.2. Kontingenstabeller.

I dette afsnit vil vi først definere hvad kontingenstabeller er samt notationen for disse. Derefter viser vi ud fra gennemgangen af et eksempel, hvordan analysen af en 2-dimensonal kontingenstabbel foregår.

3.2.1. Katagorier og niveauer.

Hvis man ønsker at sammenholde f.eks. hudkræfts placering på kroppen i forhold til arten af hudkræft for et antal hudkræft-patienter, kan man opskrive en kontingenstabell. I det eksempel vi her vil vise, og som er refereret i Dobson (1983), har man to kategorier (variable), nemlig sygdommens art og sygdommens placering. Disse to katagorier er igen inddelt i niveauer nemlig fire sygdomsarter og tre placeringer af sygdommen. Dette giver anledning til en 4x3-kontingenstabell som vist i tabel 3.1.

værtypel placering	hoved, hals	krop	arme, ben	total
fregner	22	2	10	34
spredt overfladisk pigmentering	16	54	115	185
knuder	19	33	73	125
ubestemmelig	11	17	28	56
total	68	106	226	400

Tabel 3.1. De observerede værdier. (Dobson 1983).

Fordi denne tabel kun indeholder to kategorier, kaldes den 2-dimensonal. Det er altså antallet af kategorier, der bestemmer kontingenstabellens dimension. Antallet af niveauer indenfor en kategori bestemmer størrelsen af I,J,K..., hvor man betegner størrelsen af den 2-dimensionale tabel som (IxJ), den 3-dimensionale som (IxJxK) osv.. I ovennævnte eksempel er I=4 og J=3, og vi har derfor en (4x3) 2-dimensiontal tabel.

Tabellen tilføjes endvidere en række samt en søjle, der indeholder summen af hhv. søjlerne og rækkerne. Disse kaldes søjlemarginaler hhv. rækmemarginaler.

3.2.2. Indeksering af kontingenstabeller.

Notationen for kontingenstabeller er som almindelig matrix notation, dvs. man bruger række-nummeret som første indeks og søjlenummeret som andet indeks. Pladserne i tabellen kaldes celler, og det antal, der står i en celle, betegnes n_{ij} og det totale antal er N. Herunder følger en "abstrakt" 2-dimensonal kontingenstabel.

		Kategori B							
		B_1	B_j	B_J	totaler (marginaler)		
Kate- gori A	A_1	n_{11}		n_{1j}		n_{1J}	$n_{1\cdot}$		
	.								
	.								
	.								
	A_i	n_{i1}		n_{ij}		n_{iJ}	$n_{i\cdot}$		
	.								
	.								
	.								
	.								
	A_I	n_{I1}		n_{Ij}		n_{IJ}	$n_{I\cdot}$		
	totaler (marg.)	$n_{\cdot 1}$		$n_{\cdot j}$		$n_{\cdot J}$	$N=n_{\cdot \cdot}$		

Figur 3.1.

Betegnelsen "..." bruges, når der summeres over en række hhv. søjle. I ovenstående figur betyder $n_{1\cdot}$, således, at man i første række summerer over j, dvs.

$$n_{1\cdot} = n_{11} + n_{12} + \dots + n_{1j} \cdot$$

I visse tilfælde kan det være mere hensigtmæssigt at betragte cellesandsynligheder istedet for celleantal, og tilsvarende at betragte marginalsandsynligheder. En marginalsandsynlig-

hed $\hat{p}_{.j}$ er da lig med

$$\frac{n_{.j}}{N} \text{ eller } \hat{p}_{i.} = \frac{n_{i.}}{N}$$

og cellesandsynligheden

$$\hat{p}_{ij} = \frac{n_{ij}}{N} .$$

3.2.3. Analyse af en 2-dimensionalkontingenstabell.

Vi vil i det følgende afsnit vise, hvordan man analyserer en to dimensional kontingenstabell. Vi vil anvende det talmateriale, der er vist i kontingenstabellen i tabel 3.1, der iøvrigt forudsættes multinomialfordelt.

Princippet i analysen er, at man tester en hypotese H_0 , der siger, at der er total uafhængighed mellem de to variable (katagorier).

I dette eksempel vil total uafhængighed så betyde, at typen af hudkræft ikke har indflydelse på sygdommens placering på kroppen. Det betyder igen, at de marginale rækkesandsynligheder er uafhængige. Cellesandsynligheden p_{ij} kan så under H_0 udtrykkes ved

$$p_{ij} = p_{i.} \cdot p_{.j} .$$

Vi kan så estimere det forventede celleantal \hat{m}_{ij} under denne hypotese. Anvendes maximum likelihood metoden (jf. kap. 3.1), fås MLE:

$$\hat{m}_{ij} = N \cdot \hat{p}_{ij} = N \cdot \hat{p}_{i.} \cdot \hat{p}_{.j} = \frac{n_{i.} \cdot n_{.j}}{N} .$$

Er forskellen mellem \hat{m}_{ij} og n_{ij} generelt for stor, må hypotesen forkastes. Undersøgelserne foretages ved en χ^2 -test (se nedenfor).

Analysens første trin er at beregne \hat{m}_{ij} for hver celle. Ved at betragte tabellen, ses det, at der skal estimeres seks celleantal, idet der skal estimeres for tre rækker og to søjler. Den resterende række og søjle er da også fastlagt

(fittet) ud fra marginaltabellerne. Generelt vil der i en IxJ tabel, der testes for uafhængighed, være $(I-1)(J-1)$ parametre, der skal estimeres, der er altså $(I-1)(J-1)$ frihedsgrader. Tabel 3.2 viser de forventede celleantal beregnet ud fra denne metode.

vævstype \ placering	j_1	j_2	j_3	
i_1	5.8	9.0	19.2	34
i_2	31.5	49.0	104.5	185
i_3	21.2	33.2	70.6	125
i_4	9.5	14.8	31.7	56
	68	106	226	400

Tabel 3.2. De forventede værdier.

De forventede samt de observerede værdier anvendes så til beregningen af

$$\chi^2 = \sum \frac{(\text{observeret} - \text{forventet})^2}{\text{forventet}}$$

hvor der summeres over alle celler.

Fordelingen χ^2 (Pearson's χ^2 -test) vil være en tilnærmet χ^2 -fordeling med i dette tilfælde seks frihedsgrader. Det kan generelt vises, at hvis $\underline{Y} = (y_1, \dots, y_m)$ følger en m-dimensiontal multinomisk fordeling med parametrene $(N, p_1, p_2, \dots, p_m)$ og med $\underline{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$, og $p_1 = p_1(\underline{\theta}_1, \dots, \underline{\theta}_k)$, så vil

$$\begin{aligned} \chi^2 &= \frac{(y_1 - Np_1)^2}{Np_1} + \frac{(y_2 - Np_2)^2}{Np_2} + \dots + \frac{(y_m - Np_m)^2}{Np_m} \\ &= \sum_{j=1}^m \frac{(y_j - Np_j)^2}{Np_j} \end{aligned}$$

være tilnærmet χ^2 -fordelt med $m-1-k$ frihedsgrader for store N.

Vi udregner nu χ^2 og får med indsættelse i formlerne for denne, at

$$\begin{aligned}
x^2 &= \frac{(22-5,8)^2}{5,8} + \frac{(2-9)^2}{9} + \frac{(10-19,2)^2}{19,2} + \\
&\quad \frac{(16-31,5)^2}{31,6} + \frac{(54-49)^2}{49} + \frac{(115-104,5)^2}{104,5} + \\
&\quad \frac{(19-21,1)^2}{21,2} + \frac{(33-33,2)^2}{33,2} + \frac{(73-70,6)^2}{70,6} + \\
&\quad \frac{(11-9,5)^2}{9,5} + \frac{(17-14,8)^2}{14,8} + \frac{(28-31,7)^2}{31,7} \\
&= 45,2 + 5,4 + 4,4 + 7,6 + 0,5 + 1,0 + \\
&\quad 0,2 + 0 + 0,1 + 0,2 + 0,3 + 0,4 \\
x^2 &= 65,3 .
\end{aligned}$$

Antallet af frihedsgrader = $3 \cdot 2 = 6$.

I en tabel over χ^2 -fordelinger finder vi, at $\chi^2_{0,95}(6) = 12,59$, altså er værdien 65,3 signifikant for H_0 på 5% niveau. Selv på et 0,1% niveau må vi forkaste vor hypotese H_0 , idet $\chi^2_{0,999}(6) = 22,46$ og vor $\chi^2 = 65,3$ langt overstiger denne værdi.

Generelt kan det siges, at jo større x^2 -værdi man får, desto dårligere passer modellen på den hypotese, man tester.

En anden ofte anvendt teststørrelse er L^2 (også kaldet G^2 eller $-2 \log Q$), hvor

$$L^2 = -2 \sum (\text{observerede}) \log \left(\frac{\text{forventede}}{\text{observerede}} \right).$$

Hvis den valgte model er rimeligt god, vil værdierne af x^2 og L^2 være nogenlunde ens for store N, men ved få observationer kan der være ret stor forskel på dem.

3.3. Log-lineære modeller for kontingenstabeller.

Vi vil i dette afsnit beskrive hvad en log-lineær model er, og hvad den bruges til. Det skal i denne forbindelse bemærkes, at begrebet "model" her refererer til en eller anden "teori" om ssh-fordeling af observationerne, og at parametrene i modellen repræsenterer den "effekt" en bestemt faktor eller kombination af faktorerne har, i fastlæggelsen (determinering) af de værdier, observationerne antager.

Log-lineær modellerne har egentlig udgangspunkt i variansanalysen (som vi dog ikke vil komme nærmere ind på her), og har derfor "arvet" en del begreber her fra, blandt andet begrebet "vekselvirkning" som er et meget centralt begreb i log-lineær modellerne.

Vi vil i beskrivelsen af log-lineær modellerne bl.a. komme ind på første ordens vekselvirkning mellem par af variable for 2-dimensionale kontingenstabeller, dernæst anden-ordens vekselvirkning mellem tripler af variable i tredimensionale kontingenstabeller og siden generalisere de log-lineære modeller til også at gælde for n-dimensionale kontingenstabeller. Til sidst vil vi foretage en afgrænsning af log-lineær modellerne til de hierarkiske modeller.

3.3.1. Log-lineære modeller for 2-dimensionale tabeller.

Hypotesen om uafhængighed mellem to variable kan beskrives som

$$(1) \quad p_{ij} = p_i \cdot p_j$$

Denne relation beskriver en bestemt struktur eller model for et sæt data, nemlig den, at der i en population gælder at, sandsynligheden for at en observation falder i den ij'te celle af tabellen, simpelthen ^{er} lig produktet af marginal sandsynlighederne.

Det er imidlertid ofte lettere at beskæftige sig med en model hvor p_{ij} kan udtrykkes som en sum af marginal sandsynlighederne eller som en funktion af disse. Ved at tage den naturlige logaritme til (1) får man følgende formel:

$$(2) \quad \log p_{ij} = \log p_i + \log p_j$$

Dette kan med hjælp af det teoretiske antal m_{ij} , hvor

$$m_{ij} = N p_{ij} = \frac{m_i \cdot m_j}{N},$$

omskrives til:

$$(3) \quad \log m_{ij} = \log m_{i\cdot} + \log m_{\cdot j} - \log N .$$

Ved at summere over i får vi

$$(4) \quad \sum_{i=1}^I \log m_{ij} = \sum_{i=1}^I \log m_{i\cdot} + I \log m_{\cdot j} - I \log N .$$

Summeres på lignende vis over j fås

$$(5) \quad \sum_{j=1}^J \log m_{ij} = J \log m_{i\cdot} + \sum_{j=1}^J \log m_{\cdot j} - J \log N$$

Når der summeres over både i og j får man

$$(6) \quad \sum_{i=1}^I \sum_{j=1}^J \log m_{ij} = J \sum_{i=1}^I \log m_{i\cdot} + I \sum_{j=1}^J \log m_{\cdot j} - IJ \log N$$

Med analogi til variansanalysen kan ligning (3) så omskrives til

$$(7) \quad \log m_{ij} = u + u_1(i) + u_2(j)$$

hvor

$$u = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log m_{ij}$$

$$u_1(i) = \frac{1}{J} \sum_{j=1}^J \log m_{ij} - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log m_{ij}$$

$$u_2(j) = \frac{1}{I} \sum_{i=1}^I \log m_{ij} - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log m_{ij}.$$

Formel (7) er almindeligt kendt som en log-lineær model og beskriver således en lineær model for logaritmen til celleantallene. Det skal bemærkes, at man istedet for celleantallet m_{ij} kan bruge cellesandsynligheden p_{ij} i formel (7). u 'erne får da andre talværdier, men deres betydning (se nedenfor) vil stadig være den samme. Vi vil i kapitel 4 netop benytte cellesandsynligheder istedet for celleantal. u siges at repræsentere "det totale gennemsnit", dvs. gennemsnittet af ^{de}smætlige celler.

$u_1(i)$ repræsenterer "hoved effekten" af den i'te kategori af variabel 1, dvs. den betegner afvigelsen af række gennemsnittet fra det totale gennemsnit og

$u_2(j)$ betegner "hoved effekten" af den j'te kategori af

variabel 2, hvilket vil sige, afvigelsen af sjølegennemsnittet fra det totale gennemsnit. Som en konsekvens af dette må

$$\sum_{i=1}^I u_1(i) = \sum_{j=1}^J u_2(j) = 0$$

3.3.2. Første-ordens vekselvirkning.

Det der imidlertid er interessant at besæftige sig med er, når de variable ikke er uafhængige. For at vise en afhængighed mellem to variable tilføjes et ekstra led i formel(7):

$$(8) \quad \log m_{ij} = u + u_1(i) + u_2(j) + u_{12}(ij)$$

hvor de numeriske indices på parameteren betegner de variable der er afhængighed imellem og de alfabetiske indices betegner niveauerne af disse variable..

$u_{12}(ij)$ kaldes et første-ordens vekselvirkningsled og betegner således vekselvirkningen mellem niveau'erne i og j af variabel 1 og 2 hhv.. Også vekselvirkningen er fremkommet som afvigelser fra et gennemsnit og vi har da, at

$$\sum_{j=1}^J u_{12}(ij) = \sum_{i=1}^I u_{12}(ij) = 0$$

Hvis man derfor skal teste hypotesen om uafhængighed, er det det samme, som at teste om alle vekselvirknings ledende er lig nul.

Det skal bemærkes at i modellen givet ved formel(8) er de forventede værdier lig de observerede antal. Dette skyldes, at antallet af parametre er lig antallet af celler i modellen og af den grund kaldes (8) for den maksimale model for 2-dimensionale tabeller. Her er $\hat{p}_{ij} = \frac{n_{ij}}{N}$, der er således ikke noget at teste.

3.3.3. Log-lineære modeller for 3-dimensionale og multi-dimensionale tabeller.

For en 3-dimensional $I \times J \times K$ kontingenstabbel vil den maksimale log-lineære model se således ud:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{12}(ij) + u_{13}(ik) + u_{23}(jk) + u_{123}(ijk)$$

og det sidste led med tre indices kaldes for anden-ordens vekselvirkning.

En sådan model indeholder altså det totale gennemsnit, hovedeffekt parametrene for hver af de variable, første-ordensvekselvirknings parametre for hvert par af variable, samt en parameter, der repræsenterer en mulig anden-ordens vekselvirkning mellem de tre variable.

I den maksimale model vil det totale antal parametre være lig med celleantallet, idet der er I led af typen $u_1(i)$, hvilket giver $I-1$ uafhængige led, der bliver $(I-1)(J-1)$ uafhængige led af typen $u_{12}(ij)$. Tilsvarende for u_2, u_3, u_{13} og u_{23} . Endelig er der $(I-1)(J-1)(K-1)$ led af typen $u_{123}(ijk)$: Sammen med det ene led fra u giver det i alt IJK led (ses umiddelbart ved udregning) svarende til antallet af celler.

Det er herefter let at udvide log-lineær modellen til også at gælde for multidimensionale kontingenstabeller. Har man f.eks. en n -dimensional kontingenstabbel vil den tilsvarende maksimale log-lineær model indeholde følgende typer af led:

- det totale gennemsnit
 - n hovedeffekt's parametre
 - ($\frac{1}{2}$) første-ordens vekselvirkningsparametre
 - ($\frac{3}{2}$) anden-ordens vekselvirkningsparametre
 - ⋮
 - $1(n-1)$ -ordens vekselvirkningsparameter
- (Der er set bort fra de alfabetiske indices).

I forhindelse med analysen af kontingenstabeller bruger man oftest en bestemt type log-lineær modeller, nemlig de hierarkiske modeller, som vi vil beskrive i det følgende.

3.3.4. Hierarkiske modeller.

For at en model kan betegnes som en hierarkisk model gælder der, at for hvert højere-ordens led der indgår i modellen, skal også de lavere-ordens led, der kan komponeres af de variable i højere-ordens leddet, indgå. Dette betyder at, hvis f.eks. anden-ordens leddet u_{123} indgår i modellen, så skal både hoved effekt's leddene u_1, u_2, u_3 og alle første-ordens leddene indgå i modellen. Hvis et af disse led mangler, er modellen ikke en hierarkisk model.

Til at illustrere dette kan vi f.eks. se på en model hvor det er specificeret at der er uafhængighed mellem u_1 og u_2 dvs. $u_{12}=0$. For en hierarkisk model vil dette medføre at også $u_{123}=0$. Den log-lineære model for dette tilfælde vil se således ud:

$$\log m_{ijk} = u + u_1(i) + u_2(j) + u_3(k) + u_{13}(ik) + u_{23}(jk)$$

og det man umiddelbart kan læse ud af den er, at der ikke er nogen vekselvirkning mellem variabel 1 og 2 på hvert niveau af variabel 3 (1 og 2 er (betinget) uafhængige givet 3).

Man kan så diskutere det rimelige/logiske i den begrænsning, der lægges ved indførelsen af hierarkiske modeller. Det lyder dog umidelbart rimeligt, at hvis der er trefaktor (2.-ordens) vekselvirkning mellem variabel 1,2 og 3, så må der også være gensidige tofaktor (1.-ordens) vekselvirkninger. Tilfælde af denne type kan dog tænkes at forekomme, f.eks. kunne man forestille sig, at man indenfor den medicinske forskning kan finde synergistisk virkning af tre præparater samlet, medens der parvis ikke er nogen vekselvirkning, men iøvrigt vil vi ikke beskæftige os med ikke-hierarkiske modeller.

3.4. Estimation i hierarkiske log-lineære modeller.

Hvis man vil estimere i log-lineære modeller, kunne man tænke sig at estimere de relevante u 'er under den aktuelle hypotese og ud fra disse beregne de forventede celleantal. Da u 'erne angiver afvigelser fra et gennemsnit, vil man derved få et mål for, hvor godt modellen passer for hvert enkelt (vekselvirknings-) led i den log-lineære model. Normalt vil man dog ikke beregne u 'erne, men benytte dem til en mere overordnet beskrivelse af den aktuelle model.

3.4.1. Valg af metode.

Der kan anvendes flere kriterier for valget af en log-lineær model. Dette valg kan f.eks. foretages ud fra et kendskab til det bagved liggende virkelighedsområde, og valget vil afspejle hypotesen om vekselvirkning mellem de indgående kategorier. Vi introducerer en kort skrivemåde for hierarkiske modeller. Det kan illustres ved et eksempel for en 4-dimensonal tabel:

$$\log m_{ijkl} = u + u_1 + u_2 + u_3 + u_4 + u_{12} + u_{13} + u_{14} + u_{24} + u_{124}$$

Denne model kan kort og enentydigt symboliseres som:

$$[13] \quad [124],$$

idet $[124]$ viser, at leddet u_{124} er med i modellen, og da der er tale om en hierarkisk model, følger det, at leddene $u_1, u_2, u_4, u_{12}, u_{14}$ og u_{24} også må være repræsenteret. Tilsvarende viser $[13]$ at u_{13} og dermed u_3 (og u_1) er repræsenteret i modellen (u er altid med). Vi kan kalde $[13] [124]$ for modellens genererende klasse (Haberman, 1974), jvf. kap.4.

3.4.2. Principper for estimation.

I en 2-dimensonal kontigenstabbel vil der kun være to modeller, der har interesse (ved en given sandsynlighedsfordeling) nemlig modellerne [1][2] og [12]. Man vil da teste modellen [1][2], idet en forkastelse af denne model automatisk implicerer, at modellen [12] er den aktuelle (medmindre andre sandsynlighedsfordelinger kan være aktuelle).

Som vist i kap. 3.2 (jvf. kap. 3.1) findes MLE ud fra marginaltabellerne, og man laver så χ^2 -test. Som nævnt i kap. 3.3 er den maksimale model uinteressant at teste, da $\hat{m}_{ij} = n_{ij}$ (dette gælder den maksimale model for en vilkårlig dimension).

Går man videre til tabeller af højere dimensioner, viser det sig, at man i forbindelse med estimationen kan komme ud for to forskellige muligheder, nemlig at der findes eksplisitte udtryk for MLE, eller at der ikke findes eksplisitte udtryk. I sidstnævnte tilfælde må man anvende iterative metoder til estimering af de forventede celleantal. Disse metoder findes beskrevet i de fleste af de i litteraturlisten anførte bøger om emnet, f.eks Bishop m.fl. (1975).

I førstnævnte tilfælde findes MLE ud fra de marginale tabeller, principielt som i det 2-dimentionalt tilfælde, men aktuelt skal de "rigtige" marginaltabeller vælges, og dette valg afhænger af den konkrete model.

Vi viser nu, hvorledes man kan nå frem til et eksplisit udtryk for \hat{m}_{ijk} i den 3-dimensionale model [13][23]. Vi kan skrive \hat{m}_{ijk} som

$$\hat{m}_{ijk} = \exp(u + u_1(i) + u_2(j) + u_3(k) + u_{13}(ik) + u_{23}(jk))$$

Da der ikke er vekselvirkning mellem variabel 1 og 2, har vi

$$\hat{m}_{i \cdot k} = \exp(u + u_1(i) + u_3(k) + u_{13}(ik)) \sum_j \exp(u_2(j) + u_{23}(jk))$$

og $\hat{m}_{\cdot jk} = \exp(u + u_2(j) + u_3(k) + u_{23}(jk)) \sum_i \exp(u_1(i) + u_{13}(ik))$

og $\hat{m}_{\cdot \cdot k} = \exp(u + u_3(k)) \sum_i \sum_j \exp(u_1(i) + u_2(j) + u_{13}(ik) + u_{23}(jk))$

$$= \exp(u + u_{3(k)}) \sum_i \exp(u_{1(i)} + u_{13(ik)}) \sum_j (u_{2(j)} + u_{23(jk)})$$

Det ses af ovenstående, at hvis udtrykket for $m_{i..k} \cdot m_{..jk}$ divideres med udtrykket for $m_{...k}$, vil summationsfaktoren kunne forkortes væk, og vi får

$$m_{ijk} = \frac{m_{i..k} \cdot m_{..jk}}{m_{...k}}$$

og MLE for m_{ijk} bliver så

$$\hat{m}_{ijk} = \frac{n_{i..k} \cdot n_{..jk}}{n_{...k}}$$

Det er således muligt at angive et eksplisit udtryk for de forventede celleantal i denne model. Det rejser så spørgsmålet om, hvilke betingelser der skal være opfyldt for at kunne finde et eksplisit udtryk.

I kap. 4 angiver vi, at det netop er tilfældet for den type modeller, der kaldes dekomposable.

4. GRAFER, HYPERGRAFER OG KONTINGENSTABELLER.

4.1. Indledning.

I de foregående kapitler har vi introduceret grafer, hypergrafer og kontingenstabeller hver for sig. I dette kapitel vil vi prøve at sammenfatte disse begreber, så det munder ud i en anvendelse af grafer og hypergrafer som hjælpemiddel til analyse af multidimensionale kontingenstabeller. Der indføres hermed en ny klasse modeller for kontingenstabeller, grafiske modeller, der viser sig at være en delklasse af hierarkiske modeller, men indeholder dekomposable modeller som en delklasse, dvs. dekomposable modeller \subseteq grafiske modeller \subseteq hierarkiske modeller.

4.1.1. Mere om kontingenstabeller.

Vi indfører en notation for kontingenstabeller, der dels gør det muligt at behandle tabellerne uden på forhånd at angive dimensionen, dels bevirker, at overgang til sammenligning med grafer bliver umiddelbart mere overskuelig.

Mængden af kategorier kaldes C , antal kategorier $|C|$ og de enkelte kategorier betegnes γ ($\gamma \in C$). Hvert γ opdeles i kriterier(niveauer) I_γ , og mængden af celler er mængden $I = \prod_{\gamma \in C} I_\gamma$, hvor her enkelt celle kan betegnes med vektoren $i = (i_\gamma, \gamma \in C)$. Antal observationer i en celle kan betegnes $n(i)$, og cellesandsynligheden med $P(i)$. Vi vil fremover blot skrive i istedet for i . Hvis $a \subseteq C$, betegnes de dertil svarende marginale celler med i_a , og antal observerede i de marginale celler betegnes $n(i_a)$. Tilsvarende betegnes den marginale cellesandsynlighed med $P(i_a)$. Det antages, at man har en multinomialfordeling.

Som eksempel til illustration af ovenstående kan vi se på en tredimensional kontingenstabbel, dvs. $|C|=3$ og de enkelte kategorier i C kaldes R , S og T (det kunne f.eks. stå for

race, stilling og tro) vi har så $C = \{R, S, T\}$, og R, S og T er de enkelte ζ 'er. I_R , I_S og I_T er mængden af niveauer i hver af de tre kategorier. Hvis I_R indeholder J niveauer, der hver betegnes med et tal, og I_S og I_T tilsvarende indeholder K hhv. L niveauer, haves en $I \times K \times L$ kontingenstabell, og $i = (j, k, l)$ med $j=1, 2, \dots, J$, $k=1, 2, \dots, K$, og $l=1, 2, \dots, L$. $n(i)$ svarer så til den tidligere brugte n_{jkl} . Sættes $a = \{R, S\}$, kan man f.eks. skrive i som (J, K, l) , og de marginale antal $n(i_a)$ bliver ... $_1$; tilsvarende gælder for $P(i)$ og $P(i_a)$.

Den generelle log-lineære model kan i det tredimensionale tilfælde betegnes

$$\log P_{jkl} = u + u_1^j + u_2^k + u_3^l + u_{12}^{jk} + u_{13}^{jl} + u_{23}^{kl} + u_{123}^{jkl}$$

Dette kan kort skrives som

$$\log P(i) = u_a(i_a), \text{ hvor } u_a \text{ er en funktion af } i_a.$$

Hvis vi indfører, at u_\emptyset er en konstant(vektor) svarende til $a=\emptyset$, får vi netop det generelle log-lineære udtryk, og der gælder de sædvanlige bånd på de enkelte led(summerer til 0) og restriktioner på ledene indbyrdes, når der er tale om hierarkiske modeller(se kapitel 3).

Som omtalt i kapitel 3, kan hierarkisk model specifceres ved en generende klasse $\mathcal{C} = \{\mathcal{C}_i\}$, $\mathcal{C}_i \subseteq C$, hvor det for elementerne i skal gælde, at de parvis er forskellige, ikke er "delmængder af hinanden", dvs. at de er maximale. Formelt kan det skrives:

$$u_a = 0, \text{ hviss der ikke finde et } c \in \mathcal{C}, \text{ hvor } a \subseteq c.$$

Hvis f.eks. $|C|=4$ og $\mathcal{C} = \{\{1, 2\}, \{1, 3\}, \{4\}\}$ er bl.a.

$u_{124} = u_{14} = 0$ (der er mange andre u_a 'er, der er 0) i overensstemmelse med den tidligere definition af hierarkiske modeller.

4.1.2. Markovfelter.

Vi betragter nu en graf, hvor de enkelte knuder symboli-

serer kategorier (γ) fra en kontingenstabell (kanternes betydning vil fremgå efterhånden). Vi har altså grafen (V, E) , $V \subseteq C$. Man definerer så:

En ssh er Markov mht. (V, E) , hvis det gælder:

M1 $P(i) > 0$ for $\forall i \in I$

M2 $P(I_j) = \prod_{i \in I_j} P(i)$ hvis $j \notin V$, I_j er ligefordelt og uafhængig af de øvrige kategorier

M3 Hvis $r, s \in C$ og $r \neq s$, gælder $r \perp s$

hvor symbolet \perp anvendes om betinget uafhængighed: Hvis det om tre stokastiske variable X, Y og Z gælder, at X er uafhængig af Y givet Z , skrives det $X \perp Y \mid Z$. M2 svarer til, at kategorier, der ikke er repræsenteret i grafen, er irrelevante for den statistiske analyse, og man kan derfor udelade kategorier af denne art fra kontingenstabellens grafiske repræsentation; vi sætter altså $C = V$.

Det kan vises, at en ssh P er markov mht. (C, E) , hvis der findes funktioner ϕ_a , $a \subseteq C$ således, at $\phi_a = 0$ medmindre a er en komplet delmængde af C og

$$\log P(i) = \sum_{a \subseteq C} \phi_a(i)$$

Beviset udelades, men det ses, at det for de her definerede $\phi_a(i)$ gælder $\{\phi_a(i)\} \subseteq \{u_a(i_a)\}$ fra de hierarkiske modeller, og at de repræsenterer vekselvirkninger mellem (knude)naboer i grafen. Det skal bemærkes, at betegnelsen ovenfor 'a er en komplet delmængde af C' skal forstås såledet, at delgrafen induceret ved knudemængden a er komplet (men ikke nødvendigvis maximal).

Det kan endvidere vises, at følgende fire udsagn er ækvivalente:

(1) p er Markov

(2) $a \perp b$, når c er en snitmængde for a og b

(3) $\forall a \subseteq C: a \perp a^c$

(4) $\forall e \in E: e \perp e^c$

I tilfælde (2) skal det gælde, at $a \cup b \cup c = C$ og at a, b og c er disjunkte.

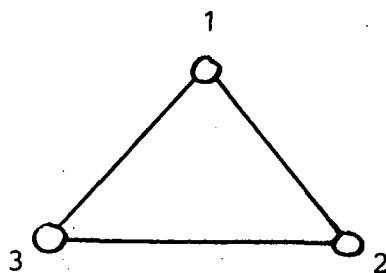
4.1.3. Sammenfatning af teorien for Markovfelter og kontingenstabeller.

Vi kan nu sammenfatte det foregående til en definition af en grafisk model:

En grafisk model er en hierarkisk model, hvis graf kan repræsenteres ved den genererende klasse hypergrafts 2-opdeling.

Et par supplerende bemærkninger vil tydeliggøre definitionen. Den genererende klasse hypergrafts 2-opdeling er en graf bestående af klier således, at enhver delgraf induceret af punktmængden hørende til en kant i hypergrafen er en klike, jvf fig. 2.20.

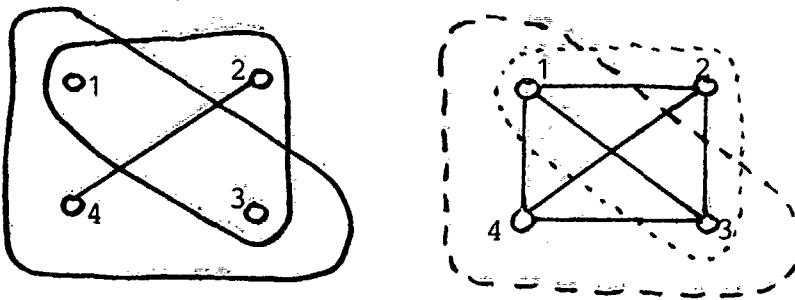
Det kunne umiddelbart virke som enhver hierarkisk model er grafisk, men et par eksempler vil vise, at det ikke er tilfældet. Den hierarkiske model med $C = \{1, 2, 3\}$ og $\mathcal{C} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ er den hierarkiske model uden 2.-ordens vekselvirkning. Grafisk skulle den se således ud:



Figur 4.1.

Umiddelbart ser grafen ud til at repræsentere en dekomposabel model; den repræsenterer den dekomposable (og dermed grafiske) model $C = \{1, 2, 3\}$, $\mathcal{C} = \{\{1, 2, 3\}\}$, men ikke modellen $C = \{1, 2, 3\}$, $\mathcal{C} = \{\{1, 2\}, \{2, 3\}, \{1, 3\}\}$.

Et andet eksempel på en hierarkisk, ikke-grafisk model er $C = \{1, 2, 3, 4\}$, $\mathcal{C} = \{\{1, 2, 3\}, \{1, 3, 4\}, \{2, 4\}\}$. G.k. hypergrafen og dens 2-opdeling hørende til denne model er vist i figur 4.2.

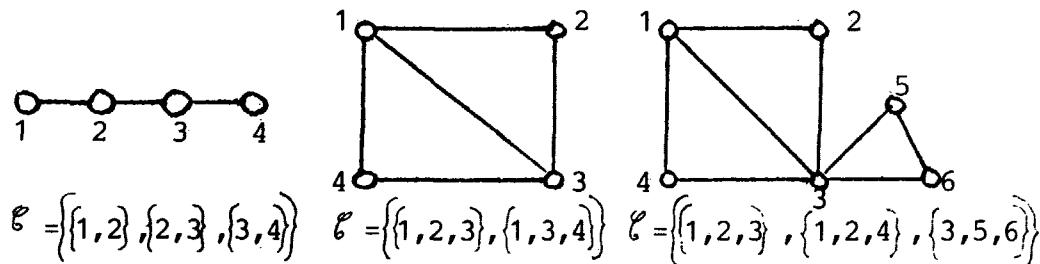


Figur 4.2.(a)+(b)

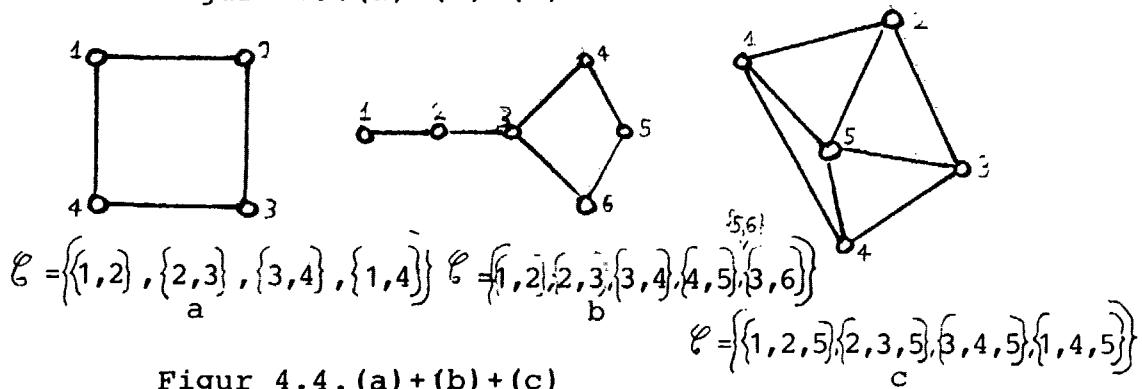
Det ses, at g.k. hypergrafens 2-opdeling er en komplet graf, dvs. hypergrafens kanter ikke er kliker i dens 2-opdeling (de er komplette delmængder, men ikke maximale).

Grafen i figur 4.2.(b) er derfor ikke en model for den nævnte hierarkiske model, men for modellen med $\mathcal{C} = \{\{1, 2, 3, 4\}\}$.

Figur 4.3(a)+(b)+(c) viser eksempler på hierarkiske modeller, der er dekomposable (og dermed grafiske) mens figur 4.4.(a)+(b)+(c) er eksempler på hierarkiske modeller, der er grafiske, men ikke dekomposable.



Figur 4.3.(a)+(b)+(c)



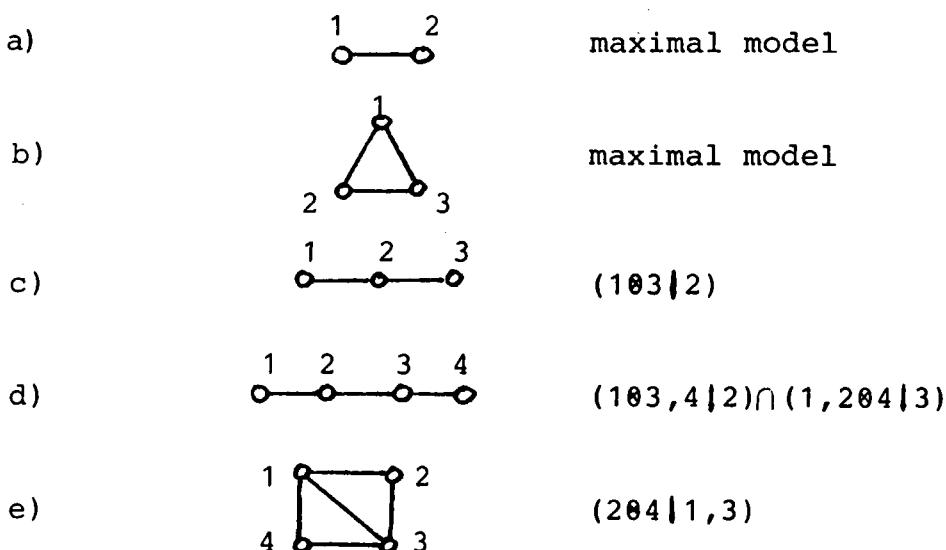
Figur 4.4.(a)+(b)+(c)

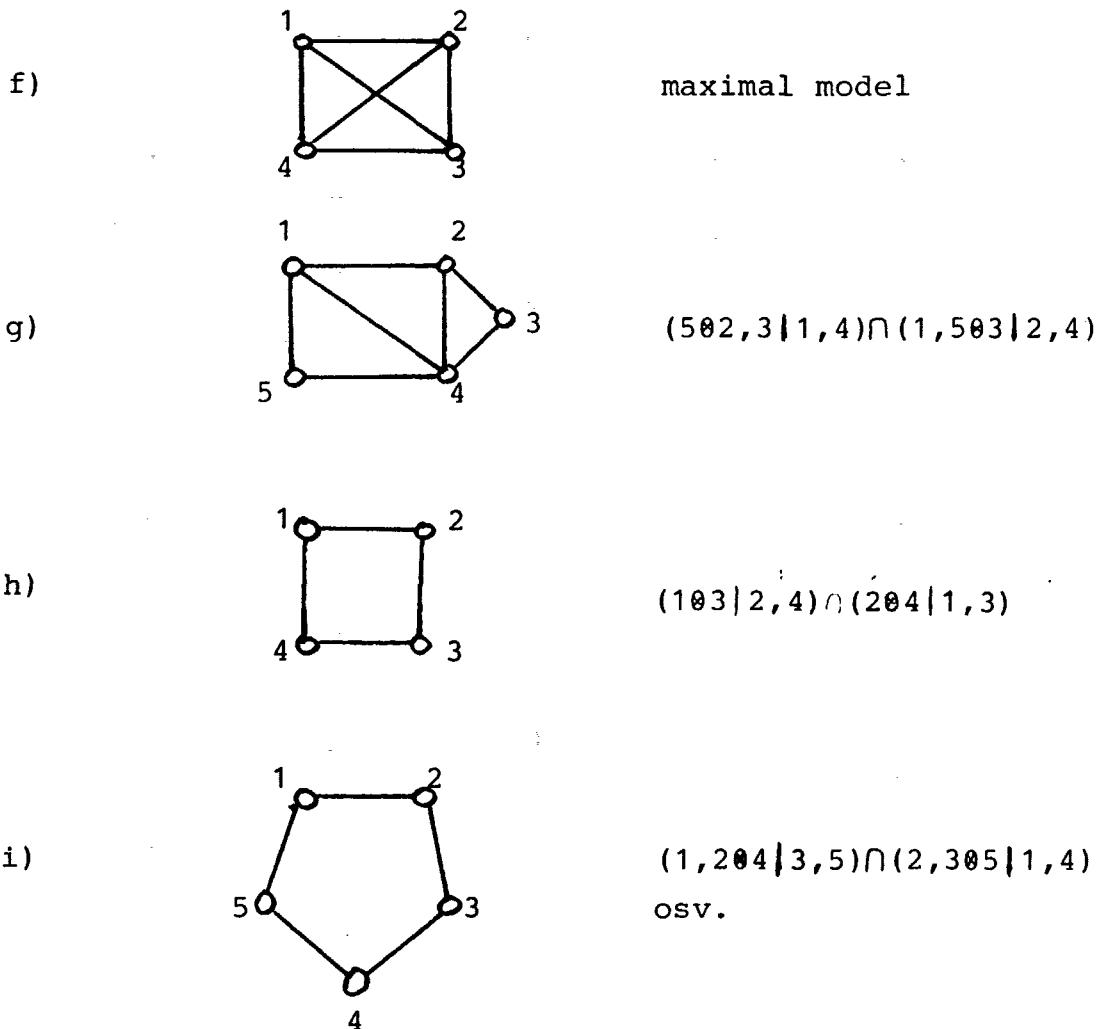
Bruger vi nu den ved Φ_a definerede Markov-egenskab, ses det, at grafiske modeller netop er de modeller, der besidder denne egenskab (udgør et Markov-felt). Det betyder, at knuderne i grafen repræsenterer hovedvirkningerne i den log-lineære model, mens grafens kanter repræsenterer 1.-ordens vekselvirkninger og klierne i grafen (kanterne i g.k. hypergrafen) repræsenterer højere-ordens vekselvirkninger.

Da Markov-egenskaben er ækvivalent med de nævnte egenskaber for betinget uafhængighed (1) og (2)-(4) ovenfor, vil de grafiske modeller umiddelbart give nogle væsentlige oplysninger, blot ved at betragte den grafiske repræsentation for disse modeller.

Det kan yderligere vises, at dekomposable modeller har den egenskab, at man kan opstille et explicit udtryk for maximum likelihood estimatet (jvf. kapitel 3.4). Det kan gøres ud fra en dekomposition af grafen, men vi vil i øvrigt ikke komme ind på metoden. Den er bl.a. beskrevet i Lauritzen (1979). Det betyder, at grafer kan give hurtigt overblik over, hvordan man mest hensigtsmæssigt kan vælge/behandle en statistisk model for kontingenstabeller.

Eksempler på brug af grafer til at finde betingede uafhængigheder gives i figur 4.5., idet skrivemåden $(1,2 \otimes 4|3)$ betegner: 1 og 2 er uafhængige af 4 givet 3.

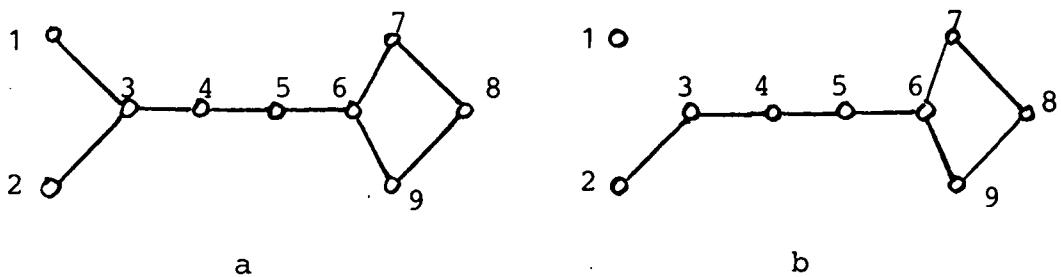




Figur 4.5.

De i figur 4.5.(h) og (i) viste grafer er ikke dekomposable, da de indeholder en kreds af længde ≥ 4 . Forskellen på dekomposable og ikke-dekomposable modeller (grafer) ligger som før nævnt i, at man for dekomposable grafer kan opstille et explicit udtryk for M.L.E., mens man for ikke-dekomposable modeller må bruge iterative metoder.

Det kan imidlertid i nogle tilfælde være hensigstmæssigt at se på grafer for ikke-dekomposable modeller, idet man dels kan se på betinget uafhængighed (jf. figur 4.5.(h)+(i)), dels kan dele af grafen være dekomposabel, og disse (denne) kan behandles som dekomposable modeller.



Figur 4.6. (a)+(b)

Grafen i figur 4.6.(a) er ikke dekomposabel, men knuderne 1,2,3,4 og 5 udgør tilsammen en dekomposabel del af grafen (modellen) og vil kunne testes som sådan, f.eks. om den kan reduceres til fig 4.6.(b).

4.2. Et eksempel på brug af grafer til test af en kontingenstabell.

Nedenstående eksempel er taget fra en artikel af Edwards og Kreiner og viser et eksempel på, hvordan en analyse af en flerdimensional kontingenstabell kan foretages. Der er i artiklen nævnt og diskuteret flere mulige metoder, der involverer grafiske såvel som ikke-grafiske metoder. Vi vil blot referere en af dem, der anvender en grafisk model.

Den mest almindelige metode til udvælgelse består af to trin. Det første trin er udvælgelse af en basismodel (start-model) og andet trin er trinvis forbedring/simplificering af modellen. Ved at sammenligne med kapitel 1., kan man se (i det følgende), at man allerede i første trin har været gennem punkterne 1.-3., som til dels er sammenfaldende, idet man f.eks. i valg af sandsynlighedsfordeling foretager både idealisering og valg af matematisk repræsentation. I det her referede eksempel (og generelt i statistiske undersøgelser) er opstillingen af modellen i snæver forstand en ständig vekselvirkning mellem idealisering og test af modellen.

På første trin udvælges en basismodel. Når man hovedsaglig er interesseret i grafiske modeller, er den maximale model her et naturligt udgangspunkt, og det antages i denne forbindelse, at der er tale om

en multinomial fordeling.

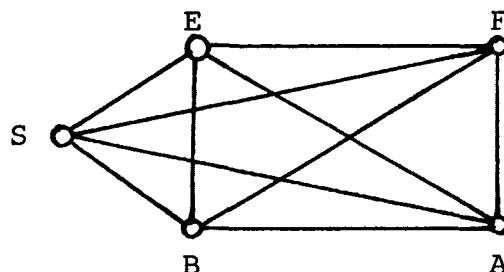
På andet trin begynder man så på selve selektionsproceduren for den grafiske model. Det, der sker, er, at man med udgangspunkt i den maximale model, tester hver kant for signifikans. Derefter fjerner man den mindst signifikante kant og undersøger de resterende kanter én for én. Atter fjernes den mindst signifikante kant, og de resterende kanter undersøges, og således fortsætter proceduren, indtil man er nået frem til en brugbar model.

Men lad os tage fat på selve eksemplet også beskrive lidt mere ud-førligt, hvad der sker undervejs.

Eksemplet omhandler en 5-dimensional kontingenstabell, der er opstillet på baggrund af materiale fra en undersøgelse foretaget af Socialforskningsinstituttet i København i 1978-1979. Undersøgelsen omfatter en gruppe mænd (arbejdere og funktionærer) i alderen 18-67 år, der blev spurgt, om de i de foregående 12 måneder havde lavet nogen form for arbejde, som de tidligere ville have betalt en håndværker for at gøre.

Formålet med undersøgelsen var bl.a. at vurdere omfanget af skatteinndragelse i bygningssektoren. De blev endvidere spurgt om alder, om de boede i hus eller lejlighed, om de boede til leje eller ejede deres bolig, og om de var faglærte, ufaglærte eller funktionærer. Materialet og betegnelsen for de variable er vist i tabel 4.1.

Den opstillede tabel er en $2 \times 3 \times 2 \times 2 \times 3$ altså en 5-dimensional kontingenstabell. Den hertil svarende maximale model [ASEBF] kan illustres ved følgende grafiske model, figur 4.7.



Figur 4.7.

F	B	E	S	A		
				Alder		
Boligform	Beskæftigelse	Ejerforhold	Svar	< 30	31-45	46-67
lejlighed	faglært	lejer	Ja	18	15	6
			Nej	15	13	9
	ufaglært	ejer	Ja	5	3	1
			Nej	1	1	1
hus	faglært	lejer	Ja	17	10	15
			Nej	34	17	19
	ufaglært	ejer	Ja	2	0	3
			Nej	3	2	0
	kontorarb.	lejer	Ja	30	23	21
			Nej	25	19	40
	ejer	Ja	8	5	1	
			Nej	4	2	2
	kontorarb.	lejer	Ja	34	10	2
			Nej	28	4	6
	ejer	Ja	56	56	35	
			Nej	12	21	8
	faglært	lejer	Ja	29	3	7
			Nej	44	13	16
	ufaglært	ejer	Ja	23	52	49
			Nej	9	31	51
	ejer	lejer	Ja	22	13	11
			Nej	25	16	12
	kontorarb.	Ja	54	191	102	
		Nej	19	76	61	

Tabel 4.1. Mænd (arbejdere el. funktionærer) krydsklassificeret efter alder, beskæftigelse, boligform, ejerforhold af bolig samt svar på spørgsmålet.

(Edwards & Kreiner, 1983).

Her er alle knuder forbundet med hinanden, dvs. modellen viser total afhængighed mellem samtlige 5 variable. Dette svarer til første trin i modelopstillingen.

På andet trin i modelopstillingen begynder så selve udvælgelsesproceduren. Første trin i denne procedure er at teste signifikansen af hver kant. Dette gøres i praksis ved, at man fjerner én kant fra basismodellen og udregner L^2 (og/eller χ^2) for den model, der bliver tilbage, når denne kant er fjernet. Det man rent faktisk tester, er, om der er betinget uafhængighed mellem de to variable, den fjernede kant forbinde givet de øvrige variable. Det gøres for modellen [ASEB], [SEBF] ved at opstille 12 todimensionale kontingenstabeller for variabel A og F, en tabel for hver kombination af niveauer for de øvrige tre variable. For hver af disse beregnes L^2 (eller χ^2), som så summeres for alle 12 tabeller, og der laves så χ^2 -test for L^2 (eller χ^2) for modellen. Antallet af frihedsgrader bliver 2·12, da hver tabel har to frihedsgrader.

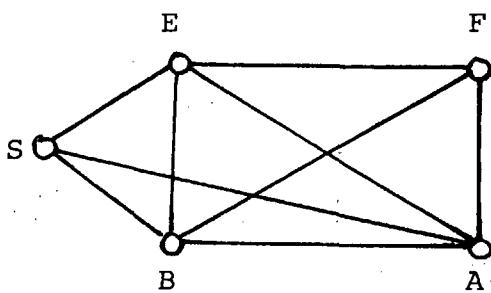
Den fjernede kant indføjes nu i modellen igen, en anden kant fjernes og den ovenfor nævnte procedure gentages. Således fortsættes indtil samtlige kanter en efter en har været fjernet fra modellen. I tabel 4.2. ser man resultatet af dette første trin i selektionsproceduren beskrevet. Modellen, der undersøges, er beskrevet i knæk-paranteserne i kolonnen 'model', og den kant, der er fjernet fra den maximale model, står i kolonnen 'kant'. 'f.g.' står for antallet af frihedsgrader, og 'p' beskriver sandsynligheden for at få en L^2 (eller χ^2)-værdi, der er større end eller lig med den fundne.

model	kant	f.g.	L^2	p	χ^2	p
[ASEB] [SEBF]	AF	24	76.09	0.0000	81.54	0.0000
[ASEB] [AEBF]	SF	18	19.81	0.3435	18.31	0.4351
[ASEB] [ASBF]	EF	18	541.68	0.0000	567.52	0.0000
[ASEB] [ASEF]	BF	24	32.37	0.1181	31.26	0.1463
[ASEF] [SEBF]	AB	32	106.36	0.0000	107.12	0.0000
[ASEF] [AEBF]	SB	24	52.98	0.0006	50.09	0.0014
[ASEF] [ASBF]	EB	24	41.06	0.0164	40.05	0.0211
[ASBF] [SEBF]	AE	24	194.32	0.0000	194.13	0.0000
[ASBF] [AEBF]	SE	18	73.71	0.0000	72.61	0.0000
[AEBF] [SEBF]	AS	24	38.77	0.0288	36.56	0.0484

A: alder, B: beskæftigelse, E: ejerforhold, F: boligform,
S: svar.

Tabel 4.2. Første trin af den første selektionsprocedure.
(Edwards & Kreiner, 1983).

Af tabel 4.2. fremgår det klart, at kanten SF er den mindst signifikante, dvs. at modellen, hvor denne kant ikke indgår, har den mindste L^2 -værdi (eller χ^2 -værdi) og den største sandsynlighed for at få en $L^2(\chi^2)$ -værdi, der er større end den fundne. At kanten SF er den mindst signifikante og derfor fjernes, medfører, at man som næste basismodel vælger modellen [ASEB] [AEBF]. Denne model kan også beskrives med betegnet uafhængighed mellem variabel S og F, givet A, E og B, formelt opskrevet som $\{S \otimes F | A, E, B\}$. Den hertil svarende grafiske model ses i figur 4.8.



Figur 4.8. Vekselvirkningsgrafen for [ASEB] [AEBF]

På andet trin i udvælgelses-proceduren tages så udgangspunkt i den i figur 4.8. beskrevne model, og resultatet af dette trin ses i tabel 4.3. Proceduren på dette trin er den samme som tidligere, hvor man fjerner kanterne en efter en og udregner L^2 for de dertil svarende modeller. I de 3 søjler yderst til højre sammenligner man modellen med den på første trin fundne nye basismodel, hvor man i de foregående 3 søjler sammenligner med den maximale model.

At beregne disse værdier for sammenligningen med den maximale model er dog overflødig, idet man blot kan lægge L^2 -værdien for modellerne i den nye basismodel til den i tabel 3 fundne L^2 -værdi for den fjernede (mindst signifikante) kant. Dermed fås L^2 -værdien for sammenligning med den maximale model. Det samme gælder for antallet af frihedsgrader; derimod er der ingen umiddelbar sammenhæng mellem p-værdierne i den ene eller den anden søjle og dette kan være begrundelsen for at medtage sammenligningen, idet man kunne forestille sig, at to kanter hvis L^2 -værdier ligger tæt, ville 'bytte om' indbyrdes ved de to sammenligninger.

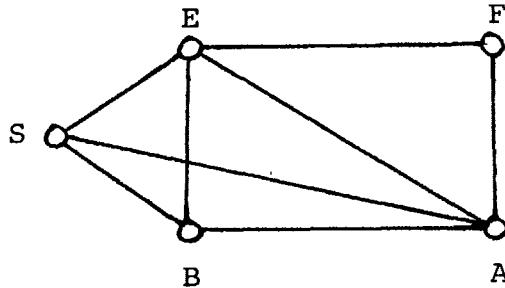
model	kant	f.g.	L^2	p	f.g.	L^2	p
[ASEB] [EBF]	AF	30	72.02	0.0000	12	57.21	0.0000
[ASEB] [ABF]	EF	27	575.75	0.0000	9	555.94	0.0000
[ASEB] [ABF]	BF	30	34.64	0.2561	12	14.83	0.2504
[ASE] [SEB] [AEF] [EBF]	AB	42	109.22	0.0000	24	89.41	0.0000
[ASE] [AEBF]	SB	30	55.25	0.0033	12	35.44	0.0004
[ASE] [ASE] [ABF] [AEF]	EB	36	51.85	0.0423	18	32.04	0.0218
[SEB] [ASB] [EBF] [ABF]	AE	36	204.19	0.0000	18	184.38	0.0000
[ASB] [AEBF]	SE	27	107.79	0.0000	9	87.98	0.0000
[SEB] [AEBF]	AS	30	39.71	0.1107	12	19.90	0.0687

Tabel 4.3. Andet trin af første selektionsprocedure.

Værdien for L^2 i modellen [ASEB] [AEF] er i den originale tabel angivet til 35.64, der så vidt vi kan se må være en trykfejl og derfor har ændret til 34.64.

(Edwards & Kreiner, 1983)

Resultatet af andet trin i selektions-proceduren viser, at kanten BF er den mindst signifikante, og dermed at både S og B er betinget uafhængige af F givet A og E, $\{S, B \otimes F | A, E\}$. Som følge heraf vil modellen [ASEB] [AEF] blive valgt som næste basismodel. Den grafiske repræsentation af denne model ses i figur 4.9.



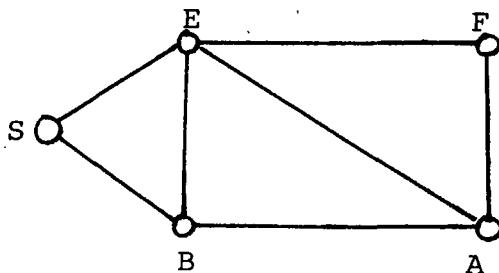
Figur 4.9. Vekselvirkningsgrafen for [ASEB] [AEF]

På tredie trin i udvælgelses-proceduren, som sker efter samme princip som på andet trin, finder man at kanten AS er den mindst signifikante og dermed at modellen [AEB] [SEB] [AEF] er den eneste model, der på dette trin kan accepteres på et 5% niveau for L^2 -testen. Resultatet af tredie trin ses i tabel 4.4, og den grafiske model for [AEB] [SEB] [AEF] ses i figur 4.10..

model	kant	f.g	L ²	p	f.g	L ²	p
[ASEB] [EF]	AF	34	88.06	0.0000	4	53.42	0.0000
[ASEB] [AF]	EF	33	594.45	0.0000	3	559.81	0.0000
[ASE] [SEB] [AEF]	AB	46	102.20	0.0000	16	85.56	0.0000
[ASE] [AEB] [AEF]	SB	42	70.07	0.0042	12	35.43	0.0004
[ASE] [ASB] [AEF]	EB	42	69.12	0.0053	12	34.48	0.0006
[ASB] [SEB] [AF] [EF]	AE	44	235.32	0.0000	14	200.68	0.0000
[ASH] [AEB] [AEF]	SE	39	122.61	0.0000	9	87.97	0.0000
[AEB] [SEB] [AEF]	AS	42	54.53	0.0931	12	19.89	0.0689

Tabel 4.4. Tredje trin af første selektionsprocedure.

(Edwards & Kreiner, 1983)



Figur 4.10. Vekselvirkningsmodellen for [AEB] [SEB] [AEF]

Alle grafiske modeller, der fremkommer af denne ved at fjerne endnu en kant, er uacceptable da $p < 0.003$. Selve udvælgelsesproceduren vil derfor standse her, og den bedste model efter denne metode, er den vist i figur 4.10.

Af denne model kan man se, at svaret på det i undersøgelsen stillede spørgsmål (vedrørende udført arbejde) kun direkte kommer an på E, den adspurgtes ejerforhold til sin bolig, og B den adspurgtes beskæftigelse. Derfor kan man, hvis man kun er interesseret i svar-variablen S, nøjes med at se på den 3-dimensionale marginaltabel SEB, da S er betinget uafhængig af A og F givet E og B. Er man derimod interesseret i flere variable vil det være tilstrækkeligt at se på de 3 3-dimensionale marginaltabeller AEB, SEB og AEF, da disse tilsammen indeholder al den statistiske information, det er muligt at uddrage af det foreliggende statistiske materiale. Man vil derfor ikke miste nogen information ved at bruge modellen frembragt af de tre genererende klasser [AEB][SEB][AEF], og man opnår en model, der med den her anvendte analyse metode, giver den bedste beskrivelse af det statistiske mate-

riale.

Det bemærkes, at enhver af de opstillede og testede modeller i udvælgelses-proceduren (naturligvis) er hierarkiske modeller med de angivne genererende klasser, og den foreliggende statistiske analyse kunne have været udført på helt tilsvarende vis ved at betragte vekselvirkningsleddene og skiftevis udelukket et af første-ordens vekselvirkningsleddene, men anvendelsen af den grafiske metode gør hele proceduren væsentligt mere overskuelig (og forståelig).

5. Opsamling og afrunding.

Inden for mange fagområder, men måske især det samfundsvidenskabelige og medicinske, anvendes kontingenstabeller af høj dimension. Vi har illustreret, hvordan man kan analysere kategoriske data v.hj.a. en hierarkisk model. Men den alt-overvejende del af rapporten omhandler begreber til forståelse af grafiske og dekomposable modeller.

Det er mange steder i rapporten omtalt, at man ved at fortolke en kontingenstabbel v.hj.a. en grafisk model reducerer antallet af modeller betragteligt. Lauritzen (1979) har opstillet en tabel over, hvor mange modeller der findes afhængig af modeltype og dimension (tabel 5.1).

type	dimension	2	3	4	5
Vekselvirkning	8	128	32768	2147483648	
Hierarkisk	5	19	167		7580
Grafisk	5	18	113		1450
Dekomposabel	5	18	110		1233

Tabel 5.1.

Som det fremgår af tabel 5.1 er det selv med en datamaskines hjælp praktisk umuligt at undersøge alle de mulige log-lineære modeller, hvis dimensionen overstiger 4. Det, man tidligere har gjort, er udfra et forhåndskendskab til virkelighedsområdet at reducere antallet af modeller. Hvis man derimod ikke har noget forhåndskendskab til problemfeltet, er den grafiske model en god førstehåndstilgang.

Man kan på en overskuelig måde reducere modellen trin for trin (ud fra den maximale model) og efterhånden nå frem til en model, der beskriver de indbyrdes sammenhænge. Metoden gør det nemt til stadighed at have overblik over den på det pågældende stadium aktuelle model, og de indbyrdes vekselvirkninger (betingede uafhængigheder) fremgår også umiddelbart af grafen. Når man så er nået frem til en statistisk set "fornuftig" model, kan den konfronteres med virkelighedsområdet, og man må så foretage en helhedsvurdering.

LITTERATURLISTE

- Andersen, E.B., Jensen, NE og Kousgaard, E:
Teoretisk statistik for økonomer. Akademisk forlag 1982.
- Berge, C: Graphs and Hypergraphs. Minieka, North-Holland 1973.
- Birkmose, M. m.fl: En undersøgelse af modellen SMEC II.
RUC rapport, matematik overbygning 1975.
- Bishop, Y., Fienberg, S & Holland, P: Discrete Multivariate Analysis. Theory and Practice. MIT Press, Cambridge Massachusetts 1975.
- Bondy, J.A & Murty, U.S.R.: Graph Theory with Applications. Macmillan Press, London 1976.
- Busacker, R. & Saaty, T.: Finite Graphs and Networks: An Introduction with Applications. McGraw-Hill, New York 1965.
- Darroch, J.N, Lauritzen, S.L & Speed, T.P.: Markov Fields and Log-linear Interaction Models for Contingency Tables. The Annals os Statistics 1980, vol 8, no 3.
- Dobson, A.: Introduction to Statistical Modelling. Chapman and Hall, London 1983.
- Edwards, D & Kreiner, S.: The Analysis of Contingency Tables by Graphical Models. England, i Biometrika (1983), 70. 3.
- Everitt, B.S.: The Analysis of Contingency Tables. Chapman and Hall, London 1977.
- Fienberg, S.: The Analysis of Cross-Classified Categorical Data. The MIT Press, Cambridge, Massachusetts, 1977.
- Haberman, S.: The Analysis of Frequency Data. The University of Chicago Press, Chicago and London. 1974.
- Harary, F.: Graph Theory. Addison-Wesley Publishing Company. Reading, Massachusetts, 1969.
- Hermann, K. & Niss, M.: Beskæftigelsesmodellen i SMEC III. En autentisk matematisk model. Nyt Nordisk Forlag. Arnold Busck, 1982.

Lauritzen, S.: Lectures on Contingency Tables. Institute of Mathematical Statistics . Univ. of Copenhagen. Sept. 1979.

Lauritzen, S, Speed, T.P. & Vijayan, K.: Decomposable Graphs and Hypergraphs. Institute of Mathematical Statistics, University of Copenhagen, June 1978. Senere udkommet i:
J. Austr. Math. Soc. 36, 12-19 (1984).

Toft, B.: En introduktion til kompleksitets- og algoritmeteori med udgangspunkt i grafteoretiske eksempler.
Odense Universitet 1984.

- 1/78 "TANKER OM EN PRAKSIS" - et matematikprojekt.
Projektrapport af Anne Jensen, Lena Lindenskov, Marianne Kesselhahn og Nicolai Lomholt.
Vejleder: Anders Madsen.
- 2/78 "OPTIMERING" - Menneskets forøgede beherskelsesmuligheder af natur og samfund.
Projektrapport af Tom J. Andersen, Tommy R. Andersen, Gert Kreinø og Peter H. Lassen.
Vejleder: Bernhelm Booss.
- 3/78 "OPGAVESAMLING", breddekursus i fysik. Nr. 3 er a jour ført i marts 1984
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 MODERNE FYSIKS HISTORIE". Nr. 5 er p.t. udgået.
Helge Kragh.
- 6/78 "NOGLE ARTIKLER OG DEBATINDLÆG OM - læreruddannelse og undervisning i fysik, og - de naturvidenskabelige fags situation efter studenteroprøret".
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-bond-graph formalismen.
Peder Voetmann Christiansen.
- 9/78 "OM PRAKSIS' INDFLYDELSE PÅ MATEMATIKKENS UDVILKING": - Motiver til Kepler's: "Nova Stereometria Doliorum Vinarioum".
Projektrapport af Lasse Rasmussen.
Vejleder: Anders Madsen.
-
- 10/79 "TERMODYNAMIK I GYMNASIET".
Projektrapport 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". Nr. 12 er udgået
Mogens Brun Heefelt
- 13/79 "CAVENDISH'S FORSØG I GYMNASIET".
Projektrapport af Gert Kreinø.
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.
Børnehelix 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 realitivity".
-
- 24/80 "MATEMATIKOPFATTELSER HOS 2.G'ERE".
a+b 1. En analyse. 2. Interviewmateriale.
Nr. 24 a+b er p.t. udgået.
Projektrapport af Jan Christensen og Knud Lindhardt Rasmussen.
Vejleder: Mogens Niss.
- 25/80 "EKSAMENSOPGAVER", Dybdemodulet/fysik 1974-79.
- 26/80 "OM MATEMATISKE MODELLER".
En projektrapport og to artikler.
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 "FUSIONENERGIEN - - - ATOMSAMFUNDETS ENDESTATION".
Oluf Danielsen.
- Nr. 30 er udgået.
Udkommer medio 1982 på Fysik-, Matematik- og Kemilærernes forlag.
- 31/80 "VIDENSKABSTEORETISCHE PROBLEMER VED UNDERVISNINGSSYSTEMER BASERET PÅ MENGDELÆRE".
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 HJELP AF MEKANISKE IMPEDANSMALINGER OG MOSSBAUER-EFFEKTMALINGER".
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-NATURVIDENSKA-BELIGE 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 HISTORIETEORI 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.
- Nr. 38 er p.t. udgået
- 39/81 "TIL KRITIKKEN AF VÆKSTØKONOMIEN".
Jens Højgaard Jensen.
- 40/81 "TELEKOMMUNIKATION I DANMARK - opлаг til en teknologivurdering".
Projektrapport af Arne Jørgensen, Bruno Petersen og Jan Vedde.
Vejleder: Per Nørgaard.
- Nr. 40 er p.t. udgået
- 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 "BARSEBÄCK 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

68/83 "STOKASTISKE MODELLER I POPULATIONSGENETIK"
- til kritikken af teoriladede modeller.

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

Vejleder: Jørgen Larsen.

69/83 "ELEVFORUDSÆTNINGER I FYSIK"
- en test i l.g med kommentarer

Albert Chr. Paulsen

70/83 "INDLÆRNINGS- OG FORMIDLINGSPROBLEMER I MATEMATIK PÅ VOKSENUNDERVISNSNIVEAU"

Projektrapport af Hanne Lisbet Andersen, Torben J. Andreasen, Svend Age Houmann, Helle Glerup Jensen, Keld Fl. Nielsen, Lene Vagn Rasmussen.

Vejleder: Klaus Grünbaum & Anders H. Madsen

71/83 "PIGER OG FYSIK"

- et problem og en udfordring for skolen?
Karin Beyer, Sussanne Bleaga, Birthe Olsen, Jette Reich & Mette Vedelsby

72/83 "VERDEN IFØLGE PEIRCE" - to metafysiske essays, om og af C.S. Peirce.

Peder Voetmann Christiansen

73/83 "EN ENERGIANALYSE AF LANDBRUG"
- økologisk contra traditionelt

ENERGY SERIES No. 9

Specialeopgave i fysik af Bent Hove Jensen

Vejleder: Bent Sørensen

74/84 "MINIATURISERING AF MIKROELEKTRONIK" - om videneskabeliggjort teknologi og nytten af at lære fysik

Projektrapport af Bodil Harder og Linda Skotak Jensen.

Vejledere: Jens Højgaard Jensen og Bent C. Jørgensen

75/84 "MATEMATIKUNDERVISNINGEN I FREMTIDENS GYMNASIUM"
- Case: Lineær programmering

Projektrapport af Morten Blomhøj, Klavs Frisdahl, Frank Mølgård Olsen

Vejledere: Mogens Brun Heefelt & Jens Bjørneboe

76/84 "KERNEKRAFT I DANMARK?" - Et høringsvar indkaldt af miljøministeriet, med kritik af miljøstyrelsens rapporter af 15. marts 1984.

ENERGY SERIES No. 10

Af Niels Boye Olsen og Bent Sørensen

77/84 "POLITISKE INDEKS - FUP ELLER FAKTA?"
Opinionsundersøgelser belyst ved statistiske modeller

Projektrapport af Svend Age Houmann, Keld Nielsen, Susanne Stender

Vejledere: Jørgen Larsen & Jens Bjørneboe

- 78/84 "JÆVNSTRØMSLEDNINGSEVNE OG GITTERSTRUKTUR I AMORFT GERMANIUM"
Specialerapport af Hans Hedal, Frank C. Ludvigsen og Finn C. Physant
Vejleder: Niels Boye Olsen
- 79/84 "MATEMATIK OG ALMENDANNELSE"
Projektrapport af Henrik Coster, Mikael Wennerberg Johansen, Povl Kattler, Birgitte Lydholm og Morten Overgaard Nielsen.
Vejleder: Bernhelm Booss
- 80/84 "KURSUSMATERIALE TIL MATEMATIK B"
Mogens Brun Heefelt
- 81/84 "FREKVENSAFHÆNGIG LEDNINGSEVNE I AMORFT GERMANIUM"
Specialerapport af Jørgen Wind Petersen og Jan Christensen
Vejleder: Niels Boye Olsen
- 82/84 "MATEMATIK- OG FYSIKUNDERVISNINGEN I DET AUTOMATISEREDE SAMFUND"
Rapport fra et seminar afholdt i Hvidovre 25-27 april 1983
Red.: Jens Højgaard Jensen, Bent C. Jørgensen og Mogens Niss
- 83/84 "ON THE QUANTIFICATION OF SECURITY" nr. 83 er p.t. udgået
PEACE RESEARCH SERIES NO. 1
af Bent Sørensen
- 84/84 "NOGLE ARTIKLER OM MATEMATIK, FYSIK OG ALMENDANNELSE".
Jens Højgaard Jensen, Mogens Niss m. fl.
- 85/84 "CENTRIFUGALREGULATORER OG MATEMATIK"
Specialerapport af Per Hedegård Andersen, Carsten Holst-Jensen, Else Marie Pedersen og Erling Møller Pedersen
Vejleder: Stig Andur Pedersen
- 86/84 "SECURITY IMPLICATIONS OF ALTERNATIVE DEFENSE OPTIONS FOR WESTERN EUROPE"
PEACE RESEARCH SERIES NO. 2
af Bent Sørensen
- 87/84 "A SIMPLE MODEL OF AC HOPPING CONDUCTIVITY IN DISORDERED SOLIDS"
af Jeppe C. Dyre
- 88/84 "RISE, FALL AND RESURRECTION OF INFINITESIMALS"
af Detlef Laugwitz
- 89/84 "FJERNVARMEOPTIMERING"
af Bjarne Lillethorup & Jacob Mørch Pedersen
- 90/84 "ENERGI I 1.G- en teori for tilrettelæggelse!"
af Albert Chr. Paulsen
-
- 91/85 "KVANTETEORI FOR GYMNASIET"
1. Lærervejledning
Projektrapport af: Birger Lundgren, Henning Sten Hansen og John Johansson
Vejleder: Torsten Meyer

92/85 "KVANTETEOREM FOR GYMNASIET
2. Materiale

Projektrapport af: Birger Lundgren, Henning
Sten Hansen og John Johansson

Vejleder: Torsten Meyer

93/85 "THE SEMIOTICS OF QUANTUM-NON-LOCALITY"

af Peder Voetmann Christiansen

94/85 "TREENIGHEDEN BOURBAKI - generalen, matematikeren
og ånden"

Projektrapport af: Morten Blomhøj, Klavs Frisdahl
og Frank M. Olsen

Vejleder: Mogens Niss

95/85 "AN ALTERNATIV DEFENSE PLAN FOR WESTERN EUROPE"

Peace research series no. 3

af Bent Sørensen

96/85 "ASPEKTER VED KRAFTVARMEFORSYNING"

af Bjarne Lillethorup

Vejleder: Bent Sørensen

97/85 "ON THE PHYSICS OF A.C. HOPPING CONDUCTIVITY"

Jeppe C. Dyre

98/85 "VALGMULIGHEDER I INFORMATIONSALDEREN"

af Bent Sørensen

99/85 "Der er langt fra Q til R"

Projektrapport af: Niels Jørgensen og Mikael Klintorp
Vejleder: Andur Pedersen

100/85 "TALSYSTEMETS OPBYGNING"

af Mogens Niss

101/85 "EXTENDED MOMENTUM THEORY FOR WINDMILLS
IN PERTURBATIVE FORM"

af Ganesh Sengupta

102/85 "OPSTILLING OG ANALYSE AF MATEMATISKE MODELLER, BELYST VED
MODELLER OVER KØERS FODEROPTAGEELSE OG - OMSÆTNING!"

Projektrapport af: Lis Eilertzen, Kirsten Habekost, Lill Røn
og Susanne Stender

Vejleder: Klaus Grünbaum

103/85 "ØDSLE KOLDKRIGERE & VIDENSKABENS LYSE IDEER"

Projektrapport af: Niels Ole Dam og Kurt Jensen

Vejleder: Bent Sørensen

104/85 "ANALOGREGNEMASKINEN OG LORENZLIGNINGER"

af: Jens Jæger

105/85 "THE FREQUENCY DEPENDENCE OF THE SPECIFIC HEAT AT THE GLASS
TRANSITION"

af Tage Christensen

"A SIMPLE MODEL OF AC HOPPING CONDUCTIVITY"

af Jeppe C. Dyre

Contributions to the Third International Conference on the
Structure of Non-Crystalline Materials held in Grenoble
July 1985

106/85 "QUANTUM THEORY OF EXTENDED PARTICLES"

af Bent Sørensen

107/85 "ÉN MYG GØR INGEN EPIDEMI"

- flodblindhed som eksempel på matematisk modellering af et
epidemiologisk problem.

Projektrapport af: Per Hedegård Andersen, Lars Boye, Carsten

Holst Jensen, Else Marie Pedersen og Erling Møller Pedersen
Vejleder: Jesper Larsen

108/85 "APPLICATIONS AND MODELLING IN THE MATHEMATICS CURRICULUM"

- state and trends -
af Mogens Niss

109/85 "COX I STUDIETIDEN"

- Cox's regressionsmodel anvendt på studenteroplysninger
fra RUC
Projektrapport af: Mikael Wennerberg Johansen, Poul Kattler,
og Torben J. Andreasen
Vejleder: Jørgen Larsen

110/85 "PLANNING FOR SECURITY"

af Bent Sørensen

111/85 "JØRGEN RUNDT PÅ FLADE KORT"

Projektrapport af: Birgit Andresen, Beatriz Quinones og
Jimmy Staal

Vejleder: Mogens Niss

112/85 "VIDENSKABELIGGØRELSE AF DANSK TEKNOLOGISK INNOVATION

FREM TIL 1950 - BELYST VED EKSEMPLER!"

Projektrapport af: Erik Odgaard Gade, Hans Hedal, Frank
C. Ludvigsen, Annette Post Nielsen og Finn Physant.

Vejleder: Claus Bryld og Bent C. Jørgensen

113/85 "DESUSPENSION OF SPLITTING ELLIPTIC SYMBOLS II"

Af: Bernhelm Booss og Krzysztof Wojciechowski