

RAPPORTER

80/22

**EKSAKTE METODAR FOR
ANALYSE AV TOVEGSTABELLAR**

AV
ROLF AABERGE

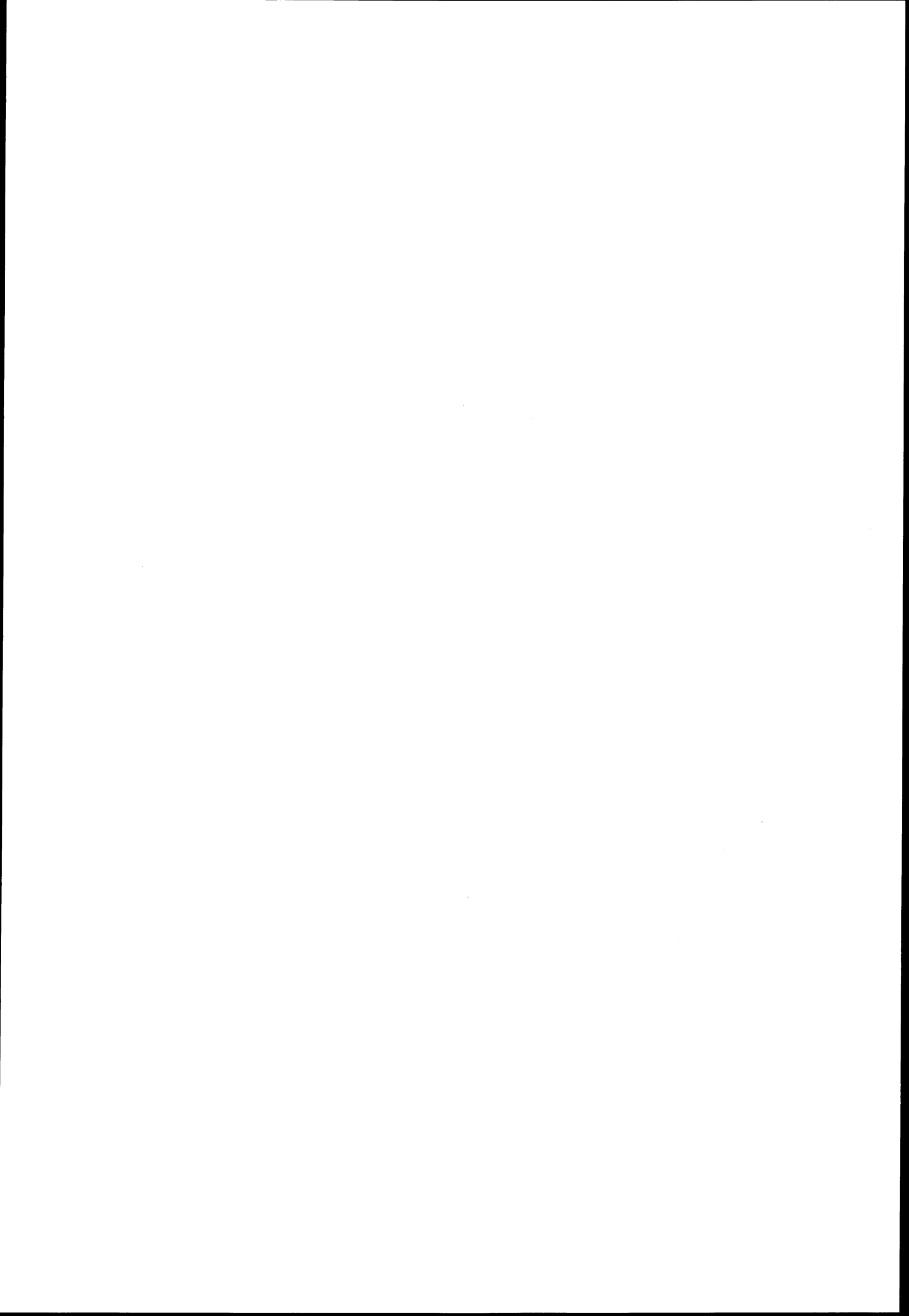
**STATISTISK SENTRALBYRÅ
OSLO**

RAPPORTER FRÅ STATISTISK SENTRALBYRÅ 80/22

EKSAKTE METODAR FOR ANALYSE
AV TOVEGSTABELLAR

AV
ROLF AABERGE

OSLO 1980
ISBN 82-537-1161-1
ISSN 0332-8422



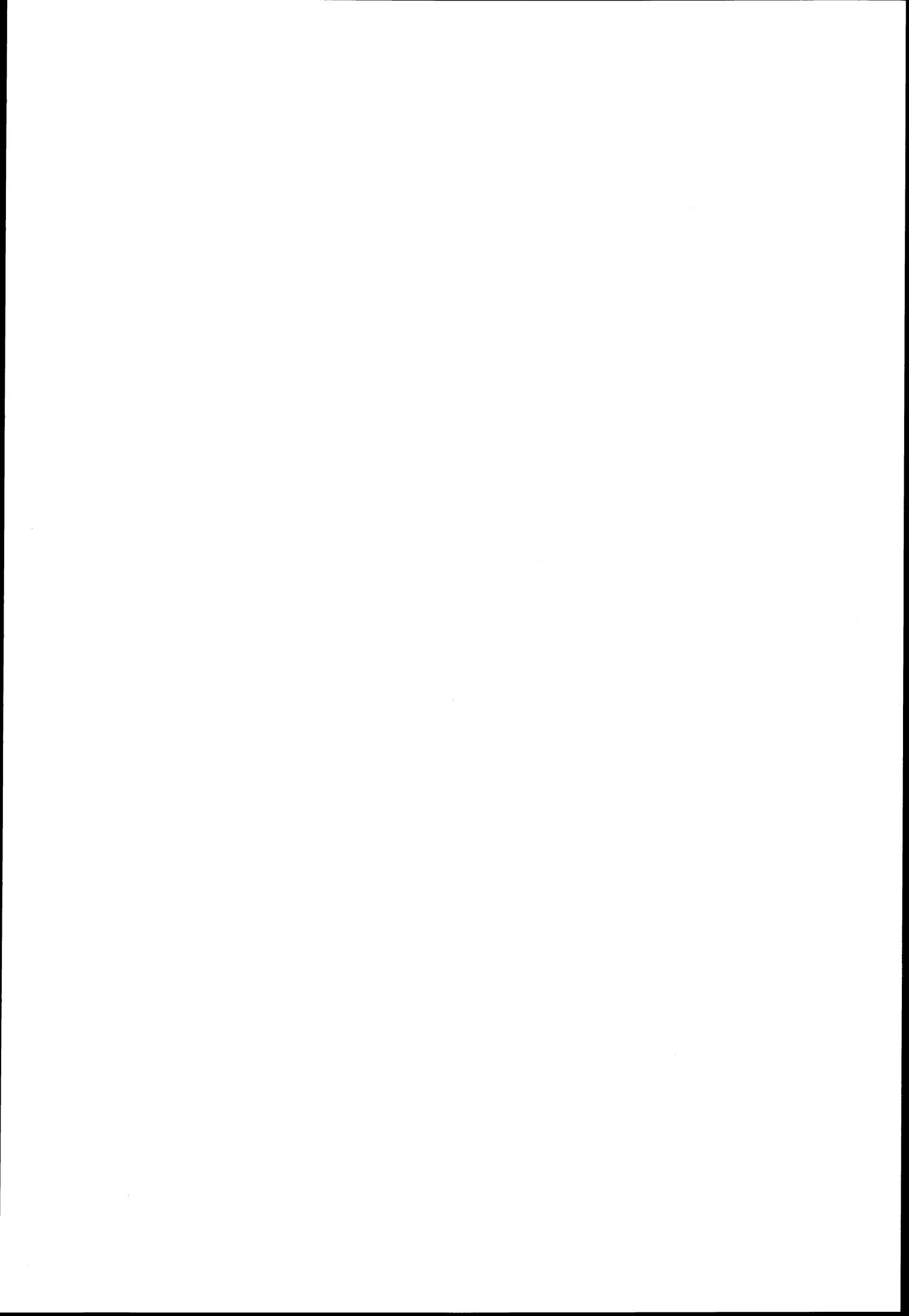
FØREORD

Bakgrunnen for arbeidet som blir presentert i denne rapporten er eit forskningsprosjekt som har vore leia av professor Erling Sverdrup. Ei av målsetjingane til dette prosjektet var å utvikle eksakte metodar for analyse av tabellar, dvs. metodar som er særskilt eigna som analysereiskap for tabellar med få observasjonar.

Resultata i denne rapporten er ei utviding og vidareføring av dei ideane som vart drøfta i Rapp 79/20.

Statistisk Sentralbyrå, Oslo, 4. juli 1980

Petter Jakob Bjerve



INNHALD

	Side
1. Innleiing	7
1.1. Motiverande eksempel	9
1.2. Definisjon av nokre viktige matematisk-statistiske omgrep	10
2. Eksistens og konstruksjon av overalt sterkaste styrkerette testar	11
2.1. Konstruksjon av tre-desisjonsreglar	15
3. Eksakte metodar for ulike hypotesar i ein to-vegstabell når modellen for tabellen er multinomisk	16
3.1. Uavhengigheit	18
3.2. Symmetri	25
3.3. Avgrensa symmetri	29
3.4. Samspelsymmetri	33
3.5. Marginal homogenitet	35
3.6. Diagonal-parameter symmetri	37
4. Eksakte metodar for ulike hypotesar i ein to-vegstabell når modellen for tabellen er produktmultinomisk	42
4.1. Homogenitet	42
4.2. Trend	46
4.3. Ulike problem i 2xs-tabellen	50
5. Asymptotiske metodar	62
6. Sentrale omgrep i to-vegstabellen uttrykt ved parametrane i den log-lineære modellen	67
Vedlegg 1:	
Konstruksjon av overalt sterkaste styrkerette testar under ein Poisson-modell	70
Vedlegg 2:	
Nullfordelinga til generaliserte Fisher-Irwins test	75
Referanser	79

-
- x) Eg vil takke Erling Sverdrup som tok initiativ til dette prosjektet, Tor Haldorsen og Jon Helgeland for mange verdifulle råd og inspirerande diskusjonar og John Dagsvik for nyttige kommentarar til manuskriptet.

S A M A N D R A G

Dei fleste matematisk-statistiske metodane som er utvikla til analyse av tabellar, byggjer på føresetnader om at talet på observasjonar i tabellcellene er "stort. Haldorsen (1977a) og (1977b) omtalar metodar som kviler på dette kravet.

I denne rapporten skal vi presentere eksakte metodar for analyse av to-vegstabellar, dvs. metodar som er gyldige sjølv om vi har små observasjonstal i tabellcellene.

1. INNLEIING

I mange undersøkingar vil observasjonane ofte gi uttrykk for kva slags kategoriar for aktuelle kjennemerke kvar av individa (einingane) høyrer til. Data av denne typen kallar vi kategoriske.

Ved å klassifisere individa frå ei undersøking med kategoriske data med omsyn på eitt eller fleire kjennemerke, får vi ein tabell som viser kor mange individ som fell under kvar av dei ulike kombinasjonane av kategoriane for kjennemerka. Dei fleste publikasjonane frå Statistisk Sentralbyrå inneheld slike tabellar. Tabellar som er resultat av ei kryssklassifisering mellom to kjennemerke blir kalla to-vegstabellar. Den enklaste av desse er 2x2-tabellen. Denne framkjem når kvart kjennemerke berre har to kategoriar.

Vi skal studere to-vegstabellar der observasjonane er genererte frå ein av følgjande tre modellar: Poisson, multinomisk eller produktmultinomisk. Val av modell er avhengig av kva slags måte data er samla inn på. For å illustrere dette skal vi gi tre enkle eksempel:

(i) Poisson-modell

Vi vil undersøke om førarar av personbil og lastebil har ulik atferd overfor stoppeplikta i vegkryss med stoppskilt. I denne situasjonen er det ikkje naturleg å fastsetje utvalstorleiken på førehand. I praksis observerer vi atferda til dei aktuelle einingane i ein avgrensa tidsperiode. Den totale utvalstorleiken blir derfor stokastisk, og modellen for tabellen blir Poisson (sjå vedlegget).

(ii) Multinomisk modell

I ei valundersøking ønskjer vi å få avslørt eventuell samanheng mellom kjønn og stemmegiving. Vi baserer undersøkinga på eit enkelt tilfeldig utval, der kun den totale utvalstorleiken (talet på personar som skal delta i undersøkinga) er fastlagt på førehand. Observasjonane frå undersøkinga kan vi presentere i ein tabell som er ei kryssklassifisering mellom kjennemerka kjønn og stemmegiving. Modellen for denne tabellen blir multinomisk (sjå kap. 3).

(iii) Produktmultinomisk modell

Vi vil undersøke om det er forskjell i ferieturaktiviteten mellom folk frå Nord-Noreg og Sør-Noreg. I denne situasjonen vil det vere naturleg

å basere undersøkinga på to enkle tilfeldige utval med kjende utvalstorleikar, eitt frå Nord-Noreg og eitt frå Sør-Noreg. Tabellen blir no ei kryssklassifisering mellom kjennemerka landsdel og ferieturaktivitet. Modellen for tabellen blir produktmultinomisk (sjå kap. 4).

I Statistisk Sentralbyrå er størstedelen av datamaterialet samla inn på ein slik måte at det er naturleg å nytte den multinomiske eller den produkt-multinomiske modellen for tabellane.

Den statistiske analysen av ein slik tabell kan gå ut på å spesifisere ulike samanhengar mellom cellesannsyna og deretter vurdere om datamaterialet som ligg føre gir grunnlag for å forkaste desse samhengane. Dei mest nytta av eksisterande analysemetodar er asymptotiske, dvs. deire eigenskapar kviler på føresetnaden om "store" utval. Vi kan derfor miste kontrollen med feilsannsynet til slike metodar når utvala er "små".

I samband med prosjektet "Analyse og estimering av kategoriske data" som har vore leia av professor Erling Sverdrup, har vi arbeidd med delprosjektet "Eksakt analyse av tabellar". Siktemålet med dette prosjektet har vore å utvikle eksakte metodar for analyseføremål i tabellar. Slike metodar er gyldige uansett tala på observasjonar i tabellcellene. I dette prosjektet har vi utvikla metoderesultat for fleirvegstabellar. Resultata i denne rapporten gjeld for to-vegstabellar. Vi vil presentere resultatata for tabellar med høgre dimensjonar i ein framtidig rapport. Ved å lese Aaberge (1979), som er ei særskilt handsaming av 2x2-tabellar, skulle det bli lettare å tileigne seg dei ideane som blir omtalt i denne rapporten.

Dei mest sentrale omgrepa i tilknytning til analyse av to-vegstabellar er uavhengigheit, symmetri, homogenitet og marginal homogenitet. Desse omgrepa har ein presis matematisk definisjon uttrykt ved ulike funksjonar av parametrane i modellen for tabellen. I rapporten viser vi kva slags informasjon desse ulike parameterfunksjonane kan gi oss. For å vurdere denne informasjonen har vi utvikla metodar med visse optimale eigenskapar.

I denne rapporten presenterer vi det analytiske verktøyet som er etablert for to-vegstabellar under prosjektet "Eksakt analyse av tabellar". Det står framleis att ein del arbeid før dei ulike metodane kan bli tekne

i bruk. Dette arbeidet går ut på å utvikle naudsynlege reknerutinar for dei metodane som vil ha størst interesse for Statistisk Sentralbyrå. Til no er det laga program for den situasjonen som er omtala i kap. 4.3. (I) (2xs-tabell). I det motiverande eksemplet (kap. 1.1.) analyserer vi ein 2x3-tabell etter den ramma som er gitt i kap. 4.3. (I). Eksempel på bruk av dei andre metodane frå denne rapporten vil bli gitt i samband med ein programomtale av eksakte metodar. Dette vil bli gjort så snart arbeidet med å utvikle naudsynlege reknerutinar er avslutta.

Rapporten er bygt opp slik at dei ulike kapitla delvis kan lesast uavhengige av kvarandre. Kap. 2 og vedlegget krev grundige kunnskapar i matematisk statistikk, medan delar av kapitla 3, 4, 5 og 6 stiller mindre krav til lesaren sine matematisk statistiske kunnskapar.

1.1. Motiverande eksempel

Med data frå Levekårsundersøkinga 1973, gir vi i tabell 1 grunnmaterialet for arbeidstidsordninga til to pendlargrupper.

Tabell 1. Lønstakarar i grupper for pendlarstatus etter arbeidstidsordning.

	Skiftarbeid	Fast nattarbeid	Dagarbeid	Sum
Veke- og langtidspendlarar	13	0	23	36
Dagpendlarar	13	4	43	60
Sum	26	4	66	96

Storparten av dei tabellane som Statistisk Sentralbyrå publiserer, viser korleis einingar i ulike grupper prosentvis fordeler seg med omsyn på ein variabel. I ein publikasjon frå Statistisk Sentralbyrå ville derfor materialet i tabell 1 bli presentert som vist i tabell 2.

Tabell 2. Lønstakarar i grupper for pendlarstatus etter arbeidstidsordning. Prosent.

Pendlarstatus	I alt	Skiftarbeid	Fast nattarbeid	Dagarbeid	Talet på personar
Veke- og langtidspendlarar .	100	36	0	64	36
Dagpendlarar	100	22	7	72	60

Denne tabellen er tidlegare publisert i Iversen (1977).

Vi vel å sjå på problemet med å kommentere tabellen som eit val mellom tre avgjerder (desisjonar):

- (i) Større førekomst av skiftarbeid og samstundes mindre førekomst av fast nattarbeid blant veke- og langtidspendlarane.
- (ii) Større førekomst av skiftarbeid og samstundes mindre førekomst av fast nattarbeid blant dagpendlarane.
- (iii) Ikkje seie noko.

Dette blir kalla tre-desisjonsproblem og er omtala i kap. 2.1. For å etablere ein desisjonsregel for dette problemet må vi bestemme ein modell for tabellen. I tabell 2 blir kvar av gruppene (i forspalten) skildra av multinomiske fordelingar, dvs. modellen for tabellen er produktmultinomisk (kap. 4). I eksempel 1 (kap. 4.3.) viser vi korleis problemet kan bli uttrykt ved parametrane i den produktmultinomiske modellen for tabellen. Som desisjonsregel skal vi derfor gjere bruk av den eksakte metoden som blir presentert i kap. 4.3. (I). Vi krev at desisjonsregelen skal ha nivå 0.05, dvs. sannsyna for å påstå (i) når (ii) er sann og (ii) når (i) er sann skal vere høgst 0.05.

For materialet i tabell 1 gir den eksakte desisjonsregelen avgjerd (i), dvs. vi kan påstå at førekomsten av skiftarbeid er større blant veke- og langtidspendlarane enn blant dagpendlarane og samstundes at det motsette er tilfelle når det gjeld fast nattarbeid.

1.2. Definisjon av nokre viktige matematisk-statistiske omgrep

La X ha fordeling P , der $P \in \mathcal{P}$. \mathcal{P} er ein eksponensiell fordelingsklasse om vi for einkvar $P \in \mathcal{P}$ har at

$$(I) \quad dP = a(P) \exp(\theta(P)Z(x)) dP_0$$

der P_0 er eit element i \mathcal{P} , θ vektor med s element og $Z = (Z_1, \dots, Z_s)'$.

La $\Theta = \theta(\mathcal{P})$, der \mathcal{P} er eksponensiell. Dersom avbildinga

$$\psi: P \rightarrow \theta(P), P \in \mathcal{P}$$

er ein-eintydig så seier vi at parametriseringa ψ^{-1} er kanonisk og den er minimal kanonisk viss dimensjonen til Θ er lik dimensjonen til \mathcal{P} .

Observatorane Z i (I) blir kalla minimal kanoniske viss parametriseringa av \mathcal{P} er minimal kanonisk.

La \mathcal{P} vere ein eksponensiell fordelingsklasse der Z er ein minimal kanonisk observator for \mathcal{P} . La P_0 vere eit element i \mathcal{P} . Da er fordelingsklassen $\tilde{\mathcal{P}}$ generert av (P_0, Z) ikkje avhengig av P_0 og Z og vidare er $\mathcal{P} \subset \tilde{\mathcal{P}}$. Vi seier at \mathcal{P} er kanonisk viss $\mathcal{P} = \tilde{\mathcal{P}}$.

Ein eksponensiell fordelingsklasse er regulær viss \mathcal{P} er kanonisk og viss vi for ein eller annan minimal kanonisk representasjon $\{P_\theta: \theta \in \Theta\}$ har at mengda Θ er open.

Θ er ei affin mengd viss $\sum_{i=1}^2 a_i \theta_i \in \Theta$ for alle $\theta_i \in \Theta$, $i = 1, 2$ og alle reelle a_1 og a_2 slik at $\sum_{i=1}^2 a_i = 1$.

2. EKSISTENS OG KONSTRUKSJON AV OVERALT STERKASTE STYRKERETTE TESTAR

I dette kapitlet føreset vi at to-vegstabellen er resultat av ei multinomisk forsøksrekke, dvs. vi har N uavhengige forsøk (utvalsstorleiken N er fast). Dersom vi i forsøket har halde marginalane til det eine kjennemerket fast, blir den underliggjande modellen produktmultinomisk. Denne situasjonen er omtala i kap. 4.

Vi har N uavhengige forsøk som kan kryssklassifiserast m.o.p. to kjennemerke A og B . A og B har respektive r og s kategoriar. Ved kvart forsøk vil ei og berre ei av hendingane $A_i \cap B_j$, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ inntreffe. I dei N forsøka er hendinga $A_i \cap B_j$ observert X_{ij} ganger. X_{ij} blir derfor talet på observasjonar i celle (i, j) til $r \times s$ -tabellen. I innleiinga av kap. 3 har vi omtalt denne tabellen meir detaljert.

La $p_{ij} = \Pr(A_i \cap B_j)$, $0 < p_{ij} < 1$ for alle i og j . Simultanfordelinga (punktsannsynet) til X_{ij} -ane er da

$$(2.1) \quad \Pr\left(\bigcap_{i=1}^r \bigcap_{j=1}^s (X_{ij} = x_{ij})\right) = \frac{N!}{\prod_{i=1}^r \prod_{j=1}^s (x_{ij})!} \prod_{i=1}^r \prod_{j=1}^s p_{ij}^{x_{ij}}$$

der

$$\sum_{i=1}^r \sum_{j=1}^s p_{ij} = 1 \quad \text{og} \quad \sum_{i=1}^r \sum_{j=1}^s X_{ij} = N.$$

La \mathcal{P} vere fordelingsklassen definert ved (2.1). Vi har at \mathcal{P} er eksponensiell sidan vi for einkvar $P \in \mathcal{P}$ har

$$(2.2) \quad dP = \left[(hp_{rs})^N \exp \left(\sum_{(i,j) \neq (r,s)} \sum x_{ij} \log \frac{p_{ij}}{p_{rs}} \right) \right] dP_0$$

der P_0 er sannsynfordelinga svarande til $p_{ij} = \frac{1}{h}$, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ og $h = rs$.

La

$$\Theta = \theta(\mathcal{P}), \theta = (\theta_{11}, \dots, \theta_{1s}, \dots)'$$

der

$$(2.3) \quad \theta_{ij} = \log \frac{p_{ij}}{p_{rs}}, \quad (i,j) \neq (r,s)$$

som er ekvivalent med

$$p_{ij} = \frac{\exp(\theta_{ij})}{1 + \sum_{(i,j) \neq (r,s)} \exp(\theta_{ij})}, \quad (i,j) \neq (r,s).$$

Følgjeleg er transformasjonen $P \rightarrow \theta(P)$, $P \in \hat{\mathcal{P}}$ definert ved (2.3) ein-ein-tydig. Vidare er $\dim \Theta = \dim \hat{\mathcal{P}} = h-1$. Dette gir at parametriseringa

$$\{P_\theta: \theta \in \Theta\}$$

gitt ved

$$(2.4) \quad dP_\theta = \left[a(\theta) \exp \left(\sum_{(i,j) \neq (r,s)} \sum \theta_{ij} x_{ij} \right) \right] dP_0$$

$$\text{der} \quad a(\theta) = h^N \left(1 + \sum_{(i,j) \neq (r,s)} \exp(\theta_{ij}) \right)^{-N},$$

er minimal kanonisk representasjon av $\hat{\mathcal{P}}$.

I motsetning til p_{ij} kan θ_{ij} ta alle moglege verdiar på tallinja. Vidare er θ_{ij} berre definert for $(i,j) \neq (r,s)$ slik at $\Theta = \mathbb{R}^{h-1}$. Θ er altså ei open mengd, og vi har at den eksponensielle fordelingsklassen $\hat{\mathcal{P}}$ er regulær.

Nedafør vil vi avgrense oss til å studere regulære hypotesar. For å få uttrykt dette kravet eksplisitt, skal vi gjere bruk av følgjande resultat frå Barndorff-Nielssen (1970)

Lemma 1. La $\hat{\mathcal{P}}$ vere ein regulær eksponensiell fordelingsklasse og la \mathcal{P}_0 vere ein hypotese til $\hat{\mathcal{P}}$. Da er dei to følgjande utsegna ekvivalente

- (i) $\hat{\mathcal{P}}_0$ er affin
- (ii) $\hat{\mathcal{P}}_0$ er regulær.

Vi vil vidare avgrense oss til dei \mathcal{P}_0 og \mathcal{P} som er slik at hypotesen \mathcal{P}_0 kan uttrykkjast som utsegn om ein parameter, γ .

La $\mathcal{P}_{\gamma, B}$ vere ein klasse av fordelingar definert ved

$$(2.5) \quad \mathcal{P}_{\gamma, B} = \{P_\theta: B\theta = v + U(\frac{\gamma}{\xi})\}$$

der $\theta \in \mathcal{M}$, $\gamma \in \mathbb{R}$, ξ er ein vektor med dimensjon d_0 , B er ei $d_1 \times (h-1)$ -matrise med rang d_1 , $d_1 = 1, 2, \dots, h-1$; $h-1 = \dim \mathcal{M}$, U er ei $d_1 \times (d_0+1)$ -matrise med rang d_0+1 , $d_0 = 0, 1, \dots, d_1-1$, $U \neq 0$, v er ein vektor med dimensjon d_1 . Elementa til ξ er ukjende parametrar medan elementa til B , U og v er kjende reelle tal.

Vi har opplagt at $\mathcal{P}_{\gamma, B} \subseteq \mathcal{P}$.

Vi skal no teste hypotesane H_j mot dei respektive alternativa A_j , $j = 1, 2, 3, 4$:

$$H_1: \gamma \leq \gamma_0 \quad (\text{eventuelt } \gamma = \gamma_0)$$

$$A_1: \gamma > \gamma_0$$

$$H_2: \gamma \leq \gamma_1 \text{ eller } \gamma \geq \gamma_2$$

$$A_2: \gamma_1 < \gamma < \gamma_2$$

$$H_3: \gamma_1 \leq \gamma \leq \gamma_2$$

$$A_3: \gamma < \gamma_1 \text{ eller } \gamma > \gamma_2$$

$$H_4: \gamma = \gamma_0$$

$$A_4: \gamma \neq \gamma_0$$

La $\mathcal{M}_B = \{\theta: B\theta = U(\frac{\gamma}{\xi}) + v\}$ der θ , γ , ξ , U og v er som i (2.5). Da får vi at

$$B(a_1\theta_1 + a_2\theta_2) = a_1B\theta_1 + a_2B\theta_2 = a_1(U(\frac{\gamma}{\xi}) + v) + a_2(U(\frac{\gamma}{\xi}) + v) = U(\frac{\gamma}{\xi}) + v$$

for alle reelle tal a_1 og a_2 som er slik at $a_1 + a_2 = 1$ og for alle $\theta_1, \theta_2 \in \mathcal{M}_B$.

Følgjeleg har vi vist at \mathcal{M}_B og dermed $\mathcal{P}_{\gamma, B}$ er affin. Sidan \mathcal{P} definert ved (2.4) er regulær, gir derfor Lemma 1 at $\mathcal{P}_{\gamma, B}$ er regulær.

Vi kan da nytte Teorem 3 (kap. 4) Lehman (1959) for å etablere følgjande resultat:

Setning 1. La hypotesen $\mathcal{P}_0 = \mathcal{P}_{\gamma, B}$, $\gamma \in H_j$, $j = 1, 2, 3, 4$ der $\mathcal{P}_{\gamma, B}$ er gitt ved (2.5). Da eksisterer det overalt sterkaste styrkerette testar med nivå ϵ for å teste $\mathcal{P}_0 = \mathcal{P}_{\gamma, B}$, $\gamma \in H_j$ mot respektive $\mathcal{P}_{\gamma, B}$, $\gamma \in A_j$; $j=1, 2, 3, 4$.

Vi skal no spesielt ta for oss testproblem av typen H_1, A_1 . Under klassen av modellar definert ved (2.5) vil vi etablere metodar for å teste $H: \gamma = 0$ mot $A: \gamma > 0$. Vi får da at fordelingsklassen $\mathcal{P}_{\gamma, B}$ må vere på formen

$$(2.6) \quad dP_{\theta} = a(\theta) \exp(\gamma(\theta)T(x) + \sum_{i=1}^{h-d-1} \tau_i(\theta)Y_i(x)) dP_0$$

der $x = (x_{11}, \dots, x_{rs})'$ og $d = d_1 - d_0$.

Funksjonsformene til γ , τ_1 , τ_2 , \dots , τ_{h-d-1} og dermed til T , Y_1 , Y_2 , \dots , Y_{h-d-1} avhenger av spesifiseringa av B , U og v .

Ved å nytte Setning 1 og Teorem 3 (kap. 4) Lehmann (1959), får vi følgjande resultat:

Setning 2. La X ha fordeling gitt ved (2.1) og la $\mathcal{P}_{\gamma, B}$ vere definert ved (2.5). For å teste $H: \mathcal{P}_{0, B}$ mot $A: \mathcal{P}_{\gamma, B}$, $\gamma > 0$ så eksisterer det ein overalt sterkast styrkerett (O.A.S.S.) test δ med nivå ε . Testen er definert ved

$$\delta(x) = \begin{cases} 1 & \text{når } T(x) > k(y_1, y_2, \dots, y_{h-d-1}) \\ \mu & \text{når } T(x) = k(y_1, y_2, \dots, y_{h-d-1}) \\ 0 & \text{når } T(x) < k(y_1, y_2, \dots, y_{h-d-1}) \end{cases}$$

der k og μ er gitt ved $E_H(\delta(X) | y_1, y_2, \dots, y_{h-d-1}) = \varepsilon$.

Setning 1 seier at for alle parametrar γ gitt ved (2.5) eksisterer det overalt sterkaste styrkerette testar for å teste problem av typen H_j mot A_j ; $j = 1, 2, 3, 4$. For å gjere bruk av Setning 1 og Setning 2 må vi derfor undersøke om hypotesen \mathcal{P}_0 kan skrivast på formen $\mathcal{P}_{0, B} = \{P_{\theta}: B\theta = v + U(\xi)\}$, der v , B og U er som i (2.5). Vi etablerer så ein O.A.S.S. test for hypotesen $\mathcal{P}_{0, B}$ under apriori-rom som er av ein dimensjon høgre enn dimensjonen til parameterrommet under hypotesen. Under hypotesen har vi dimensjon $(h-d-1)$, der h er talet på celler i tabellen og $d = \text{rang}(B) - \text{dim}(\xi)$. Vi studerer derfor modellar som har dimensjon $(h-d)$. Den fulle modellen har dimensjon $(h-1)$, slik at apriori reduksjonen er på $(d-1)$ fridomsgrader.

Merk: Når $d = 1$ har vi ingen apriori-restriksjon på parameterrommet og dermed blir $\mathcal{P}_{\gamma, B} = \mathcal{P}$.

Vi blir kvitt plageparametrane τ i sannsynfordelinga (2.6) hvis $d = h - 1$. Setning 2 gjeld framleis, men med den endring at testen no blir den overalt sterkaste blant testar med nivå ε .

Dei fleste hypotesane i kapitla 3 og 4 er presenterte på ei slik form at $d_0 = 0$, dvs. U er ein vektor med dimensjon d_1 . I kap. 3.6 studerer vi ein hypotese relativt til ein modell som er gitt på forma (2.5) med d_0 ulik 0.

2.1. Konstruksjon av tre-desisjonsreglar

Den klassiske hypoteseprøvingsteorien omtalar metodar for val mellom to avgjerder; enten å forkaste eller ikkje seie noko. For mange situasjonar i samband med tabellanalyse, vil det vere naturleg å formulere problema som tre-desisjonsproblem. Som i avsnittet ovafor skal vi studere parameteren γ under klassen av modellar gitt ved (2.5). Vi ønskjer å avgjere når $\gamma < 0$ eller $\gamma > 0$, dvs. vi har val mellom tre avgjerder (desisjonar):

$$D_1 = \text{påstå } \gamma < 0, D_2 = \text{påstå } \gamma > 0, D_3 = \text{ikkje seie noko.}$$

La observasjonsmaterialet X ha fordeling P , der $P \in \mathcal{P}$ er gitt ved (2.1). Vi definerer den statistiske metode ψ ved

$$(2.7) \quad \psi_i(X) = \text{betinga sannsynet for å velje } D_i \text{ gitt } X,$$

$$\text{der} \quad \sum_{i=1}^3 \psi_i(X) = 1.$$

Vi skal leggje følgjande krav på metoden:

(i) Nivået skal vere ϵ , dvs. sannsyna for feilaktig å påstå $\gamma < 0$ og feilaktig å påstå $\gamma > 0$ skal vere høgst ϵ .

$$E_P \psi_1(X) \leq \epsilon \quad \text{når } \gamma > 0$$

$$E_P \psi_2(X) \leq \epsilon \quad \text{når } \gamma < 0$$

(ii) Prestasjonsrettheit, dvs. sannsyna for rett utsagn skal vere minst ϵ .

$$E_P \psi_1(X) \geq \epsilon \quad \text{når } \gamma < 0$$

$$E_P \psi_2(X) \geq \epsilon \quad \text{når } \gamma > 0.$$

Blant alle ψ som tilfredsstillar (i) og (ii), ønskjer vi å finne ein ψ som maksimerer $E_P \psi_1(X)$ for $\gamma < 0$ og $E_P \psi_2(X)$ for $\gamma > 0$.

Dersom vi ser bort i frå kravet $\sum_{i=1}^3 \psi_i = 1$, har vi i røynda to separate to-desisjonsproblem om å finne to overalt sterkaste styrkerette testar ψ_1 og ψ_2 . I denne situasjonen blir $\psi_1 + \psi_2 \leq 1$. Vi kan derfor nytte Setning 2 som gir opphav til følgjande resultat:

Setning 3. La X ha fordeling gitt ved (2.1) og la γ vere definert ved (2.5).
Desisjonsregelen som går ut på å påstå

$$\gamma < 0 \text{ når } T(x) < k_1(y_1, y_2, \dots, y_{h-d-1}),$$

$$\gamma > 0 \text{ når } T(x) > k_2(y_1, y_2, \dots, y_{h-d-1})$$

og vidare påstå

$$\gamma < 0 \text{ med sannsyn } \mu_1$$

når

$$T(x) = k_1(y_1, y_2, \dots, y_{h-d-1}),$$

$$\gamma > 0 \text{ med sannsyn } \mu_2$$

når

$$T(x) = k_2(y_1, y_2, \dots, y_{h-d-1})$$

er desisjonsregelen som uniformt maksimerer prestasjonsfunksjonen ($E_p \psi$) blant alle prestasjonsrette desisjonsreglar med nivå ϵ .

I kapitla 3 og 4 presenterer vi to-desisjonsmetodar (hypotesetestar) for analyse av ulike problem i to-vegstabellen. Ved å bruke Setning 3 får vi umiddelbart fram analoge tre-desisjonsreglar. I praksis vil det ofte vere naturleg å nytte tre-desisjonsformuleringa av dei ulike problema i kapitla 3 og 4 og dermed tre-desisjonsregelen som analysereiskap (sjå kap. 1.1.).

Tre-desisjonsproblemet sett i relasjon til klassisk hypoteseprøving er utførleg drøfta i Sverdrup (1976) og (1977).

3. EKSakte metodar for ulike hypotesar i ein to-vegstabell, når modellen for tabellen er multinomisk

Bishop et al. (1975) drøftar asymptotiske metodar for ei rekkje omgrep (hypotesar) i tilknytning til analyse av to-dimensjonale tabellar. Desse omgrepa er formulerte som funksjonar av parametrane til modellen for tabellen. Nedafor skal vi presentere eksakte metodar for analyse av uavhengigheit og ulike typar symmetri.

Vi har N uavhengige forsøk som kan kryssklassifiserast m.o.p. to kjennemerke A og B. A og B har respektive r og s kategoriar. Resultatet av N slike forsøk kan vi gi i rxs-tabellen

	B_1	B_2	B_j	B_s	Sum
A_1	X_{11}	X_{12}		X_{1j}		X_{1s}	X_{1+}
A_2	X_{21}	X_{22}		X_{2j}		X_{2s}	X_{2+}
⋮							
A_i	X_{i1}	X_{i2}		X_{ij}		X_{is}	X_{i+}
⋮							
A_r	X_{r1}	X_{r2}		X_{rj}		X_{rs}	X_{r+}
Sum	X_{+1}	X_{+2}		X_{+j}		X_{+s}	N

der X -ane er multinomisk fordelt med underliggjande parametrar

	B_1	B_2	B_j	B_s	Sum
A_1	P_{11}	P_{12}		P_{1j}		P_{1s}	P_{1+}
A_2	P_{21}	P_{22}		P_{2j}		P_{2s}	P_{2+}
⋮							
A_i	P_{i1}	P_{i2}		P_{ij}		P_{is}	P_{i+}
⋮							
A_r	P_{r1}	P_{r2}		P_{rj}		P_{rs}	P_{r+}
Sum	P_{+1}	P_{+2}		P_{+j}		P_{+s}	1

der $p_{ij} = \Pr(A_i \cap B_j)$ og $\sum_{i=1}^r \sum_{j=1}^s p_{ij} = 1$.

Når situasjonen er som ovafor, seier vi at modellen for tabellen er multinomisk.

Marginalsannsyna p_{i+} og p_{+j} er definert ved

$$p_{i+} = \sum_{j=1}^s p_{ij}, \quad i = 1, 2, \dots, r$$

$$p_{+j} = \sum_{i=1}^r p_{ij}, \quad j = 1, 2, \dots, s.$$

Desse uttrykkjer sannsyna for at eit individ skal høyre til kategoriane i og j for respektive kjennemerke A og kjennemerke B .

Marginalsummane X_{i+} og X_{+j} er definert ved

$$X_{i+} = \sum_{j=1}^s X_{ij}, \quad i = 1, 2, \dots, r$$

$$X_{+j} = \sum_{i=1}^r X_{ij}, \quad j = 1, 2, \dots, s.$$

der

$$\sum_{i=1}^r X_{i+} = \sum_{j=1}^s X_{+j} = \sum_{i=1}^r \sum_{j=1}^s X_{ij} = N.$$

Vi skal no nytte resultatata frå kap. 2 til å etablere metodar for ulike funksjonar av parametrane i modellen for rxs-tabellen. Vi vil dessutan gi ein omtale av kva slags informasjon desse ulike parameterfunksjonane kan gi oss.

3.1. Uavhengigheit

Dersom $\Pr(A_i \cap B_j) = \Pr(A_i) \cdot \Pr(B_j)$ for alle i, j dvs.

$$(3.1) \quad p_{ij} = p_{i+} \cdot p_{+j}, \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, s$$

seier vi at variablane A og B er stokastisk uavhengige.

Kryssproduktforholda α_{ij} er definert ved

$$(3.2) \quad \alpha_{ij} = \frac{p_{ij} \cdot p_{rs}}{p_{is} \cdot p_{rj}} = \frac{p_{ij}/p_{is}}{p_{rj}/p_{rs}}, \quad i = 1, 2, \dots, r-1; j = 1, 2, \dots, s-1.$$

Av dette ser vi at når $\alpha_{ij} > 1$ kan vi slå fast at førekomsten av B_j relativt til B_s er større under A_i enn under A_r . Når $\alpha_{ij} = 1$ for alle i og j er A og B stokastisk uavhengige. Dette kan lett visast.

Sidan

$$\alpha_{ij} = \frac{(a_i \cdot b_j \cdot p_{ij})(a_r \cdot b_s \cdot p_{rs})}{(a_i \cdot b_s \cdot p_{is})(a_r \cdot b_j \cdot p_{rj})}, \quad i = 1, 2, \dots, r-1; j = 1, 2, \dots, s-1$$

har vi at α_{ij} er invariant under transformasjonen

$$(3.3) \quad p_{ij} \rightarrow a_i \cdot b_j \cdot p_{ij}$$

for ei kvar mengd av dei positive tala $\{a_i\}$, $\{b_j\}$ som er slik at

$\sum_{i=1}^r \sum_{j=1}^s a_i \cdot b_j \cdot p_{ij} = \sum_{i=1}^r \sum_{j=1}^s p_{ij} = 1$. Dette vil seie at kryssproduktforholda α_{ij} er mål for samspel som ikkje blandar marginalinformasjon med samspelinformasjon. Mange av dei mest kjende måla for samspel har ikkje denne eigenskapen.

Elles viser vi til Altham (1970) som gir ei utførleg grunngiving for at mål for samspel i to-vegstabellar bør vere funksjonar av kryssproduktforholda α_{ij} .

Vi skal no vise at vi kan skrive modellen for uavhengigheit på formen

$\mathcal{P}_{O,B}$

Ved å nytte (2.3) får vi for α_{ij} uttrykt ved dei kanoniske parametrene,

$$\alpha_{ij} = \exp(\theta_{ij} - \theta_{is} - \theta_{rj}), \quad i = 1, 2, \dots, r-1; j = 1, 2, \dots, s-1.$$

La $\theta = (\theta_{11}, \dots, \theta_{1s}, \theta_{21}, \dots, \theta_{2s}, \dots, \theta_{r-1,1}, \dots, \theta_{r-1,s}, \theta_{r1}, \dots, \theta_{rs-1})'$

og

$$(3.4) \quad B = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & \dots & s & \dots & s(r-2)+1 & \dots & s(r-1)+1 & \dots & h-1 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ s-1 \\ \vdots \\ s(r-2)+1 \\ \vdots \\ d_1 \end{matrix} & \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & -1 & 0 & \dots & 0 & 0 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & -1 & 0 & \dots & 0 & 0 & 0 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & -1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 1 & 0 & \dots & 0 & -1 & -1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 1 & -1 & 0 & 0 & 0 & \dots & -1 \end{pmatrix} \end{matrix}$$

der $h = rs$ og $d_1 = (r-1)(s-1)$.

Da har vi at

$$B\theta = 0$$

er ekvivalent med

$$\alpha_{ij} = 1, \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1.$$

Dette vil seie at hypotesen (modellen) for uavhengigheit kan skrivast på formen $\mathcal{P}_{0,B}$.

Vi skal no etablere ein test for hypotesen $\mathcal{P}_{0,B}$ mot alternativa $\mathcal{P}_{\gamma,B}$ (definert ved (2.5)), der B er gitt ved (3.4), $v = 0$,

$$U = (a_{11}, \dots, a_{1s-1}, a_{21}, \dots, a_{2s-1}, \dots, a_{r-11}, \dots, a_{r-1s-1})',$$

og a -ane er faste reelle tal, dvs. vi skal studere parameteren γ under modellen $\mathcal{P}_{\gamma,B}$.

A priori restriksjonane $B\theta = U\gamma$ kan dermed uttrykkjast ved

$$(3.5) \quad \theta_{ij} - \theta_{is} - \theta_{rj} = a_{ij}\gamma, \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1.$$

som igjen er ekvivalent med

$$\alpha_{ij} = \frac{p_{ij} \cdot p_{rs}}{p_{is} \cdot p_{rj}} = \exp(a_{ij}\gamma), \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1.$$

For fordelingsklassen \mathcal{P} definert ved (2.4) har vi

$$(3.6) \quad dP_{\theta} = a(\theta) \exp\left(\sum_{i=1}^r \sum_{j=1}^s \theta_{ij} x_{ij}\right) dP_0$$

(i,j) ≠ (r,s)

der

$$a(\theta) = (rs)^N \left(1 + \sum_{i=1}^r \sum_{j=1}^s \exp \theta_{ij}\right)^{-N}$$

(i,j) ≠ (r,s)

Sidan $\mathcal{P}_{\gamma, B}$ er \mathcal{P} avgrensa ved (3.5), finn vi ved å setje (3.5) inn i (3.6) at $\mathcal{P}_{\gamma, B}$ er gitt ved

$$(3.7) \quad dP_{\theta} = a(\theta) \exp\left(\gamma \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} + \sum_{i=1}^{r-1} \theta_{is} x_{i+} + \sum_{j=1}^{s-1} \theta_{rj} x_{+j}\right) dP_0.$$

Ved å nytte Setning 2 finn vi at testen som forkastar $H : \gamma = 0$ (uavhengigheit) mot $A : \gamma > 0$ når

$$Y = \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} X_{ij} > k(X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1})$$

er O.A.S.S. test

med nivå ϵ .

k vert fastsett slik at

$$\sum_{j>k} g_1(j | x_{1+}, \dots, x_{r-1+}, x_{+1}, \dots, x_{+s-1}) + \mu_1 g_1(k | x_{1+}, \dots, x_{r-1+}, x_{+1}, \dots, x_{+s-1}) = \epsilon$$

der $0 \leq \mu_1 \leq 1$,

$$g_1(y | x_{1+}, \dots, x_{r-1+}, x_{+1}, \dots, x_{+s-1}) = \sum_{x \in S_1} f(x_{11}, \dots, x_{rs})$$

$$= \sum_{x \in S_1} \frac{\binom{x_{1+}}{x_{11}, \dots, x_{1s}} \binom{x_{2+}}{x_{21}, \dots, x_{2s}} \dots \binom{x_{r+}}{x_{r1}, \dots, x_{rs}}}{N \binom{x_{+1}, \dots, x_{+s}}{x_{+1}, \dots, x_{+s}}},$$

$$S_1 = \{x: \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} = y\}, \quad x = (x_{11}, x_{12}, \dots, x_{1s}, \dots, x_{r1}, x_{r2}, \dots, x_{rs})'$$

$$x_{rj} = x_{+j} - \sum_{i=1}^{r-1} x_{ij}, \quad j = 1, 2, \dots, s, \quad x_{is} = x_{i+} - \sum_{j=1}^{s-1} x_{ij}, \quad i = 1, 2, \dots, r$$

og $f(\cdot)$ er den betingta fordelinga for (X_{11}, \dots) gitt (X_{1+}, \dots) .

Sidan den betinga fordelinga for $(X_{11}, \dots, X_{r-1, s-1})$ gitt $(X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1})$ under H (uavhengigheit) blir generalisert multivariat hypergeometrisk, er det naturleg å kalle testen for generalisert Fisher-Irwins test.

Dette resultatet kan grunnrivast ved eit kombinatorisk resonnement. Nemnaren i uttrykket for den generaliserte multivariate hypergeometriske fordelinga er lik talet på måtar vi kan dele populasjonen inn i s grupper (etter kjennemerket B) med storleikar $x_{+1}, x_{+2}, \dots, x_{+s}$. Teljaren er lik talet på måtar vi kan dele dei x_{1+} einingane i første kategori (A_1) til kjennemerket A inn i grupper med storleikar $x_{11}, x_{12}, \dots, x_{1s}$, dei x_{2+} einingane i andre kategori (A_2) inn i grupper av storleikar $x_{21}, x_{22}, \dots, x_{2s}$, osv.

Merk: Ved å ta i bruk Lemma 2 (sjå Vedlegg 2) kan vi også finne nullfordelings til testen på tradisjonelt vis.

Når vi i praksis skal gjere bruk av metoden som er omtala ovafor, ser vi bort frå randomiseringa. Vi forkastar hypotesen om uavhengigheit ($\gamma=0$) mot $\gamma > 0$, dersom dei observerte X_{i+} , $i = 1, 2, \dots, r-1$, X_{+j} , $j = 1, 2, \dots, s-1$ og X_{ij} , $i = 1, 2, \dots, r-1$; $j = 1, 2, \dots, s-1$ er slik at

$$G_1(\tilde{Y} | X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1}) = \sum_{\ell \leq \tilde{Y}} g_1(\ell | X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1}) > 1 - \epsilon.$$

der

$$\tilde{Y} = \max_x \left(\sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} : \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} < Y \right).$$

Dersom alle a_{ij} -ane er heiltal, og minst ein a_{ij} er lik 1,

blir

$$\tilde{Y} = Y - 1.$$

Når $r = 2$ og $s = 2$ (2x2-tabell), får vi Fisher-Irwins test.

Utrekninga av den betinga fordelinga g_1 (og dermed G_1) er reknemessig ressurskrevjande på den forma den ligg føre.

Elementær rekning viser at den multivariate hypergeometriske fordelinga kan splittast opp i eit produkt av $(r-1)(s-1)$ "hypergeometriske fordelingar". Dette resultatet er kjent frå Lancaster (1949). Den betinga fordelinga g_1 kan dermed skrivast på følgjande form

$$\begin{aligned}
 g_1(y|x_{1+}, \dots) &= \sum_{x \in S_1} \frac{\sum_{k=1}^{r-1} \sum_{\ell=1}^{s-1} x_{k\ell} \sum_{i=k+1}^r \sum_{j=\ell}^s x_{ij}}{\sum_{i=k+1}^r \sum_{j=\ell}^s x_{ij}} \\
 &= \sum_{x \in S_1} \prod_{k=1}^{r-1} \prod_{\ell=1}^{s-1} f'_{k\ell}(x_{k\ell})
 \end{aligned}$$

der

$$\sum_{i=k+1}^r x_{i\ell} = \sum_{i=k}^r x_{i\ell} - x_{k\ell},$$

dvs. g_1 kan uttrykkjast som konvolusjonen av dei betinga fordelingane $f'_{k\ell}$, $k = 1, 2, \dots, r-1$, $\ell = 1, 2, \dots, s-1$. Dette fører til følgjande enkle prosedyre for utrekninga av g_1 :

La

$$z_{k\ell} = a_{k\ell} x_{k\ell}$$

der $a_{k\ell}$ er heiltal.

La vidare

$$Z = \{0, 1, 2, \dots\}$$

og

$$f_{k\ell}(z) = \begin{cases} f'_{k\ell}(z/a_{k\ell}), & z/a_{k\ell} \text{ heiltal} \\ 0 & \text{ellers} \end{cases}$$

La $t = t(k, \ell)$ vere ordninga av indeksmengda (k, ℓ) , $k = 1, 2, \dots, r-1$, $\ell = 1, 2, \dots, s-1$ der ℓ går raskast, dvs. $t(1,1) = 1$, $t(1,2) = 2, \dots$, $t(1,s-1) = s-1$, $t(2,1) = s$, \dots , $t(r-1,s-1) = (r-1)(s-1)$

og la

$$y_t = z_t + y_{t-1}, \quad t = 2, 3, \dots, (r-1)(s-1)$$

der

$$\begin{aligned}
 y_1 &= z_1 \text{ og } y_{(r-1)(s-1)} = y = \sum_{k=1}^{r-1} \sum_{\ell=1}^{s-1} a_{k\ell} x_{k\ell} \\
 &= \sum_k \sum_\ell z_{k\ell} = \sum_{t=1}^{(r-1)(s-1)} z_t.
 \end{aligned}$$

Ved rekursjon finn vi

$$f^{(2)}(y) = \sum_z f_{11}(y-z)f_{12}(z)$$

$$f^{(t)}(y) = \sum_z f^{(t-1)}(y-z)f_t(z), \quad t = 3, 4, \dots, (r-1)(s-1)$$

og dermed

$$g_1(y | x_{1+}, \dots) = f^{((r-1)(s-1))}(y)$$

Dersom vi spesielt vel $a_{ij} = 1$ for alle i og j blir

$$Y = \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} Y_{ij}$$

og den betingte fordelinga g_1 blir redusert til den hypergeometriske fordelinga

$$\frac{\binom{\sum_{j=1}^{s-1} x_{+j}}{y} \binom{N - \sum_{j=1}^{s-1} x_{+j}}{r-1-y}}{\binom{N}{r-1}} \cdot \binom{r-1}{\sum_{i=1}^{s-1} x_{i+}}$$

Ovafor blei det vist at vi kunne studere uavhengigheit/avhengigheit i rxs-tabellen ved analyse av kryssproduktforholda α_{ij} (definert ved (3.2)). Samtidig peika vi på at α_{ij} har fornuftig tolkning, slik at vi kan oppnå interessant detaljinformasjon gjennom analyse av desse parametrane. Spesielt kan det vere aktuelt å undersøke om det er trend i kryssproduktforholda, dersom forspalten (eventuelt hovudet) i tabellen har ei naturleg ordning. Som vi skal sjå nedafor, kan problemstillingar med trend kome til uttrykk i modellen (3.5). Vi føreset at forspalten i tabellen har ei naturleg ordning.

$$\mathcal{A}_1 = \{a(i) : a(i) < 0 \text{ og } a(i) \text{ monotont veksande for } i > 0\}$$

og

$$\mathcal{A}_2 = \{a(i) : a(i) > 0 \text{ og } a(i) \text{ monotont avtakande for } i > 0\}$$

Ved å la

$$a_{ij} \in \mathcal{A}_1, \quad j = 1, 2, \dots, b$$

og

$$a_{ij} \in \mathcal{A}_2, j = b + 1, b + 2, \dots, s-1$$

i (3.5), får vi for $\gamma > 0$ at

$$\alpha_{1j} < \alpha_{2j} < \dots < \alpha_{rj}, j = 1, 2, \dots, b$$

og

$$\alpha_{1j} > \alpha_{2j} > \dots > \alpha_{rj}, j = b + 1, b + 2, \dots, s-1.$$

Vi har dermed vist at metoden ovafor (Generalisert Fisher-Irwins test, med særskilt val av a_{ij} -ane) også kan bli nytta til å vurdere trend i kryssprodukta α_{ij} , $i = 1, 2, \dots, r-1$; $j = 1, 2, \dots, s-1$.

Merk. Dei gitte reelle tala a_{ij} , $i = 1, 2, \dots, r-1$; $j = 1, 2, \dots, s-1$ som inngår i modellen $\mathcal{P}_{\gamma, B}$, skal fastsetjast på grunnlag av den apriori kunnskap vi har om fenomenet som vert undersøkt.

Metoden har to bruksområder:

(i) metode for å avsløre om det generelt er avhengigheit mellom kjennemerka i tabellen. Dersom det er feil i modellføresetnadene (nokre av a_{ij} -ane er feil valde), så fører dette til at metoden taper effektivitet. Sannsynet for feilaktig å påstå avhengigheit mellom kjennemerka i tabellen blir derimot ikkje forstyrra av feil i modellen.

(ii) sluttningsregel for dei underliggjande parametrane; dvs. metode for å undersøke om det eksisterer ei særskild spesifisert avhengigheit i tabellen. Vi kan formulere dette som tre-desisjonsproblem ved parameteren γ , og nytta tre-desisjonsregelen som svarar til testprosedyren gitt ved G_1 , dvs. vi kan påstå $\gamma > 0$ dersom

$$G_1(\tilde{Y} | X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1}) > 1 - \epsilon$$

og $\gamma < 0$ dersom

$$G_1(Y | X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1}) < \epsilon.$$

Nivået er ϵ , dvs. sannsyna for feilaktig å påstå $\gamma > 0$ og feilaktig å påstå $\gamma < 0$ er høgst ϵ .

3.2. Symmetri

Vi seier vi har symmetri i ein rxr-tabell dersom

$$p_{ij} = p_{ji}, i \neq j; i, j = 1, 2, \dots, r$$

som er ekvivalent med

$$\beta_{ij} = 1, i < j; j = 1, 2, \dots, r \quad \text{der} \quad \beta_{ij} = \frac{p_{ij}}{p_{ji}}.$$

Dersom vi har data som er resultat frå ei panelundersøking, kan kvart individ i utvalet klassifiserast etter same kriteriet på to ulike tidspunkt. Det vil da ofte vere interessant å vurdere materialet i den framkomne kvadratiske tabellen gjennom omgrepet symmetri. Symmetri vil seie at det er "fullstendig likevekt" mellom tilstandane; like mange har gått frå A til B som frå B til A. I Byrået vil det i samband med Arbeidskraftundersøkinga vere interessant å studere dei ulike overgangane mellom kategoriane sysselsett, delvis sysselsett og ikkje-sysselsett. Symmetri i tabellen vil her seie at sannsyna for tilgang og avgang frå sysselsett til delvis sysselsett og ikkje-sysselsett er parvis like, sannsyna for tilgang og avgang frå delvis sysselsett til sysselsett og ikkje-sysselsett er parvis like og sannsyna for tilgang og avgang frå ikkje-sysselsett til sysselsett og delvis sysselsett er parvis like. På tilsvarande måte kunne det vere interessant å studere pendling, kriminalitet, stemmegiving ved val, osv.

Vi har den situasjonen som er skildra innleiingsvis i kap. 3 med $s = r$, altså ein rxr-tabell. Vi vil nedafor vise at modellen for symmetri kan skrivast på formen $\mathcal{P}_{O,B}$.

Ved å nytte (2.3) får vi β_{ij} uttrykt ved dei minimalkanoniske parametranne

$$\beta_{ij} = \exp(\theta_{ij} - \theta_{ji}), \quad i < j.$$

La

$$(3.8) \quad B = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & \dots & r & r+1 & r+2 & r+3 & \dots & 2r+1 & \dots & 3r+2 & \dots & (r-1)r+1 & \dots & h-1 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ r-1 \\ r \\ r+1 \\ \vdots \\ \vdots \\ \vdots \\ d_1 \end{matrix} & \left(\begin{array}{cccccccccccccccc} 0 & 1 & 0 & 0 & \dots & 0 & -1 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & -1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & -1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & \dots & -1 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & \dots & 0 & \dots & -1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 & \dots & 0 & \dots & 1 & 0 & \dots & -1 \end{array} \right) \end{matrix}$$

der $h = r^2$ og $d_1 = \frac{1}{2}r(r-1)$.

Sidan

$$B\theta = 0$$

er ekvivalent med

$$\beta_{ij} = 1, \quad i < j.$$

kan hypotesen (modellen) for symmetri skrivast på formen

$$\mathcal{P}_{O,B}$$

Vi skal no etablere ein test for hypotesen $\mathcal{P}_{0,B}$ under modellen $\mathcal{P}_{\gamma,B}$ (definert ved (2.5)) der B er gitt ved (3.8), $v = 0$,
 $U = (a_{12}, a_{13}, \dots, a_{1r}, a_{23}, a_{24}, \dots, a_{2r}, \dots, a_{r-2r-1}, a_{r-2r}, a_{r-1r})'$
 og a_{ij} , $i < j$ er faste reelle tal, dvs. vi skal studere parameteren γ under modellen $\mathcal{P}_{\gamma,B}$.

A priori restriksjonane $B\theta = U\gamma$ kan dermed uttrykkjast ved

$$(3.9) \quad \theta_{ij} - \theta_{ji} = a_{ij}\gamma, \quad i < j$$

som er ekvivalent med

$$\beta_{ij} = \frac{p_{ij}}{p_{ji}} = \exp(a_{ij}\gamma), \quad i < j.$$

Sidan $\mathcal{P}_{\gamma,B}$ er \mathcal{P} (definert ved (3.6) med $s = r$) avgrensa ved (3.9), får vi ved å setje (3.9) inn i (3.6) at $\mathcal{P}_{\gamma,B}$ er gitt ved

$$(3.10) \quad dP_{\theta} = a(\theta) \exp(\gamma \sum_{i < j} a_{ij} x_{ij} + \sum_i \theta_{ii} x_{ii} + \sum_{i < j} \theta_{ji} (x_{ij} + x_{ji})) dP_0.$$

Setning 2 gir da at testen som forkastar

$$H : \gamma = 0 \text{ (symmetri) mot } A : \gamma > 0$$

når

$$Z = \sum_{i < j} \sum a_{ij} X_{ij} > k (W, X_{11}, X_{22}, \dots, X_{r-1r-1})$$

der $W_{ij} = X_{ij} + X_{ji}$ og $W = (W_{12}, \dots, W_{1r}, W_{23}, \dots, W_{2r}, \dots, W_{r-1r})$,
 er O.A.S.S. test med nivå ϵ .

Som vi skal sjå nedafør, er k berre avhengig av W. Derfor blir testen som forkastar

$$H : \gamma = 0 \text{ mot } A : \gamma > 0$$

når

$$Z > k(W)$$

O.A.S.S. test med nivå ϵ .

Under H har vi at

$$\begin{aligned} & \Pr\left(\bigcap_{i < j} (W_{ij} = w_{ij}), \bigcap_{i=1}^{r-1} (X_{ii} = x_{ii})\right) \\ &= \prod_{i=1}^{r-1} \frac{p_{ii}^{x_{ii}}}{p_{rr}} \prod_{i > j} \frac{p_{ij}^{w_{ij}}}{p_{rr}} \frac{N!}{p_{rr}^{r-1} \prod_{i=1}^{r-1} x_{ii}! (N - \sum_{i=1}^{r-1} x_{ii} - \sum_{i < j} w_{ij})!} \sum_{i < j} \frac{1}{\prod_{i < j} (x_{ij}! (w_{ji} x_{ij})!)} \end{aligned}$$

$$= \frac{\prod_{i=1}^{r-1} \left(\frac{p_{ii}}{p_{rr}}\right)^{x_{ii}} \prod_{i>j} \left(\frac{p_{ij}}{p_{rr}}\right)^{w_{ij}} p_{rr}^N}{\prod_{i=1}^{r-1} x_{ii}! \prod_{i<j} (N - \sum_{i=1}^{r-1} x_{ii} - \sum_{i<j} w_{ij})! \prod_{i<j} w_{ij}!} \prod_{i<j} \binom{w_{ij}}{x_{ij}} \binom{w_{ij}}{x_{ij}}$$

$$= \frac{N!}{\prod_{i=1}^{r-1} x_{ii}! \prod_{i<j} w_{ij}! (N - \sum_{i=1}^{r-1} x_{ii} - \sum_{i<j} w_{ij})!} \prod_{i=1}^{r-1} \left(\frac{p_{ii}}{p_{rr}}\right)^{x_{ii}} \prod_{i>j} \left(\frac{p_{ij}}{p_{rr}}\right)^{w_{ij}} p_{rr}^N$$

sidan $\prod_{i<j} \binom{w_{ij}}{x_{ij}} \binom{w_{ij}}{x_{ij}} = \prod_{i<j} 2^{w_{ij}} = \prod_{i>j} 2^{w_{ij}}$.

Dette gir vidare at

$$\Pr\left(\bigcap_{i<j} (X_{ij}=x_{ij}) \mid \bigcap_{i<j} (W_{ij}=w_{ij}), \bigcap_{i=1}^{r-1} (X_{ii}=x_{ii})\right)$$

$$= \frac{\Pr\left(\bigcap_{i<j} (X_{ij}=x_{ij}), \bigcap_{i<j} (W_{ij}=w_{ij}), \bigcap_{i=1}^{r-1} (X_{ii}=x_{ii})\right)}{\Pr\left(\bigcap_{i<j} (W_{ij}=w_{ij}), \bigcap_{i=1}^{r-1} (X_{ii}=x_{ii})\right)}$$

$$= \frac{N!}{\prod_{i=1}^{r-1} x_{ii}! \prod_{i<j} x_{ij}! \prod_{i>j} (w_{ij} - x_{ij})! (N - \sum_{i=1}^{r-1} x_{ii} - \sum_{i<j} w_{ij})!} \prod_{i=1}^{r-1} \left(\frac{p_{ii}}{p_{rr}}\right)^{x_{ii}} \prod_{i>j} \left(\frac{p_{ij}}{p_{rr}}\right)^{w_{ij}} p_{rr}^N$$

$$= \frac{N!}{\prod_{i=1}^{r-1} x_{ii}! \prod_{i<j} w_{ij}! (N - \sum_{i=1}^{r-1} x_{ii} - \sum_{i<j} w_{ij})!} \prod_{i=1}^{r-1} \left(\frac{p_{ii}}{p_{rr}}\right)^{x_{ii}} \prod_{i>j} \left(\frac{p_{ij}}{p_{rr}}\right)^{w_{ij}} p_{rr}^N$$

$$= \frac{\prod_{i<j} w_{ij}!}{\prod_{i<j} (x_{ij}!) (w_{ij} - x_{ij})!} \prod_{i<j} 2^{-w_{ij}} = \prod_{i<j} \binom{w_{ij}}{x_{ij}} 2^{-w_{ij}}$$

Vi har dermed vist at k er uavhengig av $X_{11}, X_{22}, \dots, X_{rr}$.

k vert fastsett slik at

$$\sum_{j>k} g_2(j|w) + \mu_2 g_2(k|w) = \epsilon$$

der

$$g_2(z|w) = \sum_{x_1 \in R_1} \prod_{i < j} \binom{w_{ij}}{x_{ij}} 2^{-w_{ij}},$$

$$x_1 = (x_{12}, x_{13}, \dots, x_{1r}, x_{23}, x_{24}, \dots, x_{2r}, \dots, x_{r-2r-1}, x_{r-2r}, x_{r-1r})',$$

$$R_1 = \{x_1 : \sum_{i < j} a_{ij} x_{ij} = z\} \quad \text{og}$$

$$0 \leq \mu_2 \leq 1.$$

I praksis forkastar vi hypotesen om symmetri, $\gamma = 0$ (mot $\gamma > 0$) dersom dei observerte X_{ij} ; $i < j$ og W_{ij} ; $i < j$ er slik at

$$G_2(\tilde{Z}|W) = \sum_{\ell \leq \tilde{Z}} g_2(\ell|W) > 1 - \epsilon$$

der

$$\tilde{Z} = \max_{x_1} (\sum_{i < j} a_{ij} x_{ij} : \sum_{i < j} a_{ij} x_{ij} < Z).$$

Dersom alle a_{ij} -ane er heiltal og minst ein a_{ij} er lik 1, blir

$$\tilde{Z} = Z - 1.$$

Når $r = 2$, fell metoden saman med den vi fann for 2x2-tabellen i Aaberge (1979).

I den betinga fordelinga g_2 ser vi at X_{ij} -ane ($i < j$) er uavhengige og binomisk fordelte ($W_{ij}, \frac{1}{2}$). Vi kan derfor rekne ut G_2 ved hjelp av program for den binomiske fordelinga og den diskrete konvolusjonen.

3.3. Avgrensa symmetri

Vi seier vi har avgrensa symmetri i ein kvadratisk tabell (rxr-tabell) dersom

$$p_{ij} = p_{ji}, \quad (i, j) \in S,$$

der $S \subseteq \{(i, j) : i \neq j\}$.

Når $S = \{(i, j) : i \neq j\}$ har vi altså full symmetri i tabellen.

Problemstillingar uttrykt ved avgrensa symmetri, vil som for symmetri vere spesielt aktuelle når vi har paneldata. Problemet kan f.eks. dreie seg om å undersøke tilgang og avgang frå ein kategori i forhold til dei andre. Eit eksempel på dette kan vere ei valundersøking, der hovudproblemet er å finne ut om avgangen frå det største partiet til kvar av dei andre partia er lik den tilsvarande tilgangen frå kvar av desse partia.

Ved å la

$$\beta_{ij} = \frac{p_{ij}}{p_{ji}}$$

får vi avgrensa symmetri dersom

$$\beta_{ij} = 1 \text{ for alle } (i,j) \in S.$$

Det følgjer opplagt at modellen for avgrensa symmetri kan skrivast på formen $\mathcal{P}_{0,B}$, der B no blir sett saman av dei linjene frå matrisa definert ved (3.8), som svarar til dei par $(i,j) \in S$; $i < j$. I denne situasjonen blir d_1 = talet på $(i,j) \in S$ for $i < j$.

Vi skal no etablere ein test for hypotesen $\mathcal{P}_{0,B}$ mot alternativa $\mathcal{P}_{\gamma,B}$ (definert ved 2.5) der B er $id_1 \times (h-1)$ -matrise gitt ved dei linjene i matrisa (3.8) som svarar til para $(i,j) \in S$ for $i < j$; $v = 0$, U er ein vektor (med dimensjon d_1) som er sett saman av elementa a_{ij} , $(i,j) \in S$, $i < j$, der a_{ij} -ane er faste reelle tal. Denne reformuleringa av problemet set oss i stand til å vurdere hypotesen om avgrensa symmetri uttrykt ved ein parameter, γ . For å få dette til har vi avgrensa oss til modellen $\mathcal{P}_{\gamma,B}$; dvs. vi har lagt apriori-restriksjonane

$$B\theta = U\gamma$$

på parameterrommet, dvs.

$$(3.11) \quad \theta_{ij} - \theta_{ji} = a_{ij}\gamma; \quad (i,j) \in S, \quad i < j$$

som er ekvivalent med

$$\beta_{ij} = \frac{p_{ij}}{p_{ji}} = \exp(a_{ij}\gamma); \quad (i,j) \in S, \quad i < j.$$

Sidan $\mathcal{P}_{\gamma,B}$ er \mathcal{P} (definert ved (3.6) med $s = r$) avgrensa ved (3.11), får vi ved å setje (3.11) inn i (3.6) at $\mathcal{P}_{\gamma,B}$ er gitt ved

$$(3.12) \quad dP_{\theta} = a(\theta) \exp\left(\gamma \sum_{\substack{(i,j) \in S \\ i < j}} \sum_{i < j} a_{ij} x_{ij} + \sum_{\substack{(i,j) \notin S \\ (i,j) \neq (r,r)}} \sum_{i < j} \theta_{ij} x_{ij}\right) \\ + \sum_{\substack{(i,j) \in S \\ i < j}} \sum_{i < j} \theta_{ji} (x_{ij} + x_{ji}) \Big| dP_0$$

Vi har gjort bruk av at

$$\sum_{\substack{(i,j) \in S \\ i > j}} \sum_{i > j} \theta_{ij} x_{ij} = \sum_{\substack{(i,j) \in S \\ i < j}} \sum_{i < j} \theta_{ji} x_{ji}.$$

La $W_{ij} = X_{ij} + X_{ji}$, $W_1 = (W_{ij} : i < j; (i,j) \in S; i, j = 1, 2, \dots, r)$,

og $X_1 = (X_{ij} : i < j; (i,j) \notin S; i, j = 1, 2, \dots, r)$.

Setning 2 gir da at testen som forkastar

$H : \gamma = 0$ (avgrensa symmetri) mot $A : \gamma > 0$

når

$$Z_1 = \sum_{\substack{(i,j) \in S \\ i < j}} \sum_{i < j} a_{ij} X_{ij} > k(W_1, X_1)$$

er O.A.S.S. test med nivå ϵ .

På tilsvarende måte som i kap. 3.2. finn vi at k berre er avhengig av W_{ij} , $i < j$, $(i,j) \in S$. Derfor blir testen som forkastar $H : \gamma = 0$ mot $A : \gamma > 0$ når

$$Z_1 > k(W_1)$$

O.A.S.S. test med nivå ϵ .

k vert fastsett slik at

$$\sum_{j > k} g_3(j|w_1) + \mu_3 g_3(k|w_1) = \epsilon$$

der

$$g_3(z_1 | w_1) = \sum_{R_2} \prod_{\substack{(i,j) \in S \\ i < j}} \binom{w_{ij}}{x_{ij}} 2^{-w_{ij}},$$

$$R_2 = \{x_{ij}, i < j; (i,j) \in S : \sum_{\substack{(i,j) \in S \\ i < j}} a_{ij} x_{ij} = z_1\}$$

og

$$0 \leq \mu_3 \leq 1.$$

I praksis forkastar vi hypotesen om avgrensa symmetri, $\gamma = 0$ (mot $\gamma > 0$) dersom dei observerte X_{ij} og W_{ij} , $i < j$, $(i,j) \in S$ er slik at

$$G_3(\tilde{Z}_1 | W_1) = \sum_{\ell \leq \tilde{Z}_1} g_3(\ell | W_1) > 1 - \epsilon$$

der

$$\tilde{Z}_1 = \max_x \left(\sum_{\substack{(i,j) \in S \\ i < j}} a_{ij} x_{ij} : \sum_{\substack{(i,j) \in S \\ i < j}} a_{ij} x_{ij} < Z_1 \right).$$

Når $S = \{(i,j) : i \neq j\}$ blir metoden identisk med den vi presenterte i kap. 3.2.

Som vi nemnde innleiingsvis, kan analysen av ein kvadratisk tabell (rxr) f.eks. dreie seg om problem hovudsakeleg knytta til den eine kategorien i tabellen. Som eksempel nemnde vi ei valundersøking, der ei gruppe menneske har gitt uttrykk for partipreferanse ved to val. Dersom det er r-1 aktuelle parti kan vi gi att materialet i ein rxr-tabell. Den første kategorien til kjennemerka er "veit-ikkje"-gruppa, medan dei resterande r-1 kategoriane er dei aktuelle partia. La oss tenkje oss at hovudinteressa er knytta til parti r (kategori r). For å studere eventuelle overgangar mellom dette partiet og dei andre partia samt "veit-ikkje"-gruppa, kan vi bruke resultatata ovafor med

$$S = \{(i,r) : i = 1, 2, \dots, r-1\}.$$

Demmed kan vi forkaste hypotesen om avgrensa symmetri,

$$\gamma = 0 \text{ (} p_{ir} = p_{ri}, i = 1, 2, \dots, r-1 \text{) mot } \gamma > 0$$

dersom

$$Z_1 = \sum_{i=1}^{r-1} a_{ir} X_{ir} > k (W_{1r}, W_{2r}, \dots, W_{r-1r})$$

der

$$W_{ir} = X_{ir} + X_{ri}, i = 1, 2, \dots, r-1.$$

3.4. Samspelsymmetri

Kryssproduktforholda α_{ij} definert ved (3.3) er mål for samspel i to-vegstabellar. Vi seier vi har samspelsymmetri i ein rxr-tabell dersom

$$\alpha_{ij} = \alpha_{ji}, \quad i \neq j; \quad i, j = 1, 2, \dots, r-1.$$

som er ekvivalent med

$$\lambda_{ij} = 1, \quad i < j < r$$

der

$$\lambda_{ij} = \frac{\alpha_{ij}}{\alpha_{ji}}, \quad i < j < r.$$

Når $r = 2$ (2x2-tabellen) blir samspelsymmetri ekvivalent med symmetri.

Ved å nytte (2.3) og (3.3) får vi for λ_{ij} uttrykt ved dei minimal-kanoniske parametrane

$$\lambda_{ij} = \exp(\theta_{ij} - \theta_{ji} - \theta_{ir} + \theta_{ri} - \theta_{rj} + \theta_{jr}), \quad i < j < r.$$

Vi har dermed at

$$\lambda_{ij} = 1, \quad i < j < r$$

er ekvivalent med

$$\theta_{ij} - \theta_{ji} - \theta_{ir} + \theta_{ri} - \theta_{rj} + \theta_{jr} = 0, \quad i < j < r.$$

slik at modellen (hypotesen) for samspel-symmetri kan skrivast på formen $\mathcal{P}_{0,B}$, der B er ei $d_1 \times (h-1)$ matrise med rang $(B) = d_1 = (r-1)(r-2)/2$ og $h = r^2$.

Vi føreset

$$(3.13) \quad \theta_{ij} - \theta_{ji} - (\theta_{ir} - \theta_{ri}) - (\theta_{rj} - \theta_{jr}) = a_{ij}\gamma, \quad i < j < r$$

der a_{ij} -ane er faste reelle tal, dvs. vi har ein apriori-modell av typen

$$B\theta = U\gamma$$

der U er ein d_1 -dimensjonal vektor med faste reelle tal.

La $\mathcal{P}_{\gamma, B}$ vere fordelingsklassen \mathcal{P} (definert ved (3.6)) avgrensa ved (3.13). Ved å setje inn for (3.13) i (3.6) får vi at $\mathcal{P}_{\gamma, B}$ er gitt ved

$$(3.14) \quad dP_{\theta} = a(\theta) \exp\left(\gamma \sum_{i < j < r} \sum_{ij} a_{ij} x_{ij} + \sum_{i < j < r} \sum_{ji} \theta_{ji} w_{ij}\right) \\ + \sum_{i=1}^{r-1} \theta_{ii} x_{ii} + \theta_{1r} t_1 + \theta_{r1} t_2 + \theta_{r-1r} t_3 + \theta_{r-1r} t_4 \\ + \sum_{k=2}^{r-2} \theta_{kr} u_k + \sum_{\ell=2}^{r-2} \theta_{r\ell} v_{\ell} \Big) dP_0,$$

der

$$w_{ij} = x_{ij} + x_{ji}, \\ t_1 = x_{1r} + \sum_{j=2}^{r-1} x_{1j}, \\ t_2 = x_{r1} - \sum_{j=2}^{r-1} x_{1j}, \\ t_3 = x_{r-1r} - \sum_{i=1}^{r-2} x_{ir-1}, \\ t_4 = x_{rr-1} - \sum_{i=1}^{r-2} x_{ir-1}, \\ u_k = x_{kr} + \sum_{j=k+1}^{r-1} x_{kj} - \sum_{i=1}^{k-1} x_{ik}$$

og

$$v_{\ell} = x_{r\ell} - \sum_{j=\ell+1}^{r-1} x_{\ell j} + \sum_{i=1}^{\ell-1} x_{i\ell}.$$

Setning 2 gir da at testen som forkastar

$$H : \gamma = 0 \text{ (samspelsymmetri) mot } A : \gamma > 0$$

når

$$Z_2 = \sum_{i < j < r} \sum_{ij} a_{ij} X_{ij} > k \text{ (} T, U, V, W, X_{11}, X_{22}, \dots, X_{r-1r-1} \text{)}$$

er O.A.S.S. test med nivå ε .

der

$$T = (T_1, \dots, T_4)', \quad U = (U_2, \dots, U_{r-2})', \quad V = (V_2, \dots, V_{r-2})',$$

$$W = (W_{ij} : i < j < r)'$$

Den betingta fordelinga til Z_2 gitt $X_{11}, X_{22}, \dots, X_{r-1r-1}, T, U, V$ og W under H har i denne situasjonen ei meir komplisert form enn i dei tre førre avsnitta. Men som tidlegare blir også denne fordelinga uavhengig av parametrane (p) i modellen for tabellen.

3.5. Marginal homogenitet

Når kjennemerka i to-vegstabellen har like kategoriar, vil det vere naturleg å samanlikne fordelingane til dei to setta av marginalsommar i tabellen. Som eksempel på slike tabellar kan vi nemne situasjonar med paneldata.

Vi seier vi har marginal homogenitet i $r \times r$ -tabellen dersom

$$p_{i+} = p_{+i}, \quad i = 1, 2, \dots, r-1.$$

Når $r = 2$ blir dette ekvivalent med symmetri. Uttrykt ved dei minimal-kanoniske parametrane θ_{ij} har vi marginal homogenitet dersom

$$\sum_{j=1}^r \exp \theta_{ij} = \sum_{j=1}^r \exp \theta_{ji}, \quad i = 1, 2, \dots, r-1$$

som er ekvivalent med

$$\log \left(\frac{\sum_{j=1}^r \exp \theta_{ij}}{\sum_{j=1}^r \exp \theta_{ji}} \right) = 0, \quad i = 1, 2, \dots, r-1,$$

dvs. modellen (hypotesen) for marginal homogenitet er ikkje på formen

$$P_{O,B}.$$

Vi har dermed vist at modellen (hypotesen) for marginal homogenitet ikkje er affin og i følgje Lemma 1 ikkje regulær. Dette hindrar oss i å nytte den same framgangsmåten som tidlegare for å etablere ein test for hypotesen om marginal homogenitet.

Bishop et al. (1975) peikar på at det er umogleg å uttrykke marginal homogenitet på ein enkel måte i den log-lineære modellformuleringa. Dei føreslår derfor å studere marginal homogenitet under apriori-føresetnaden om samspel-symmetri. Årsaka til at dei vel dette utgangspunktet er at samspelsymmetri i tillegg til marginal homogenitet blir det same som symmetri.

Vi skal derfor føresetje samspel-symmetri apriori, dvs.

$$(3.15) \quad \theta_{ij} - \theta_{ir} - \theta_{rj} = \theta_{ji} - \theta_{jr} - \theta_{ri}, \quad i < j < r.$$

Under (3.15) er hypotesen om marginal homogenitet ekvivalent med symmetri. I kap. 3.2 viste vi at hypotesen om symmetri var regulær. Dermed har vi at hypotesen om marginal homogenitet under modell (3.15) er regulær. I tillegg til (3.15) skal vi føresetje at

$$(3.16) \quad \theta_{ir} - \theta_{ri} = a_i \gamma, \quad i = 1, 2, \dots, r-1.$$

gjeld apriori.

Under modell (3.15) blir $\gamma = 0$ ekvivalent med marginal homogenitet.

La $\mathcal{P}_{\gamma, B}$ vere fordelingsklassen \mathcal{P} (definert ved (3.6)) avgrensa ved (3.15) og (3.16). Når vi set inn for (3.15) og (3.16) i (3.6) får vi at $\mathcal{P}_{\gamma, B}$ er gitt ved

$$(3.17) \quad dP_{\theta} = a(\theta) \exp\left(\gamma \sum_{i < j} (a_i - a_j) x_{ij} + \sum_{i < j} \theta_{ji} (x_{ij} + x_{ji}) + \sum_{i=1}^{r-1} \theta_{ii} x_{ii}\right) dP_{\theta}.$$

Vi ser at $\mathcal{P}_{\gamma, B}$ gitt ved (3.17) er med i fordelingsklassen (3.10).

Ved å la

$$a_{ij} = a_i - a_j, \quad a_r = 0; \quad i < j$$

i (3.10) får vi frå kap. 3.2 at testen som forkastar

$$H : \gamma = 0 \quad \text{mot} \quad A : \gamma > 0$$

når

$$Z_3 = \sum_{i < j} \sum (a_i - a_j) X_{ij} > k(W)$$

$$\text{der} \quad W = (W_{12}, \dots, W_{1r}, W_{23}, \dots, W_{2r}, \dots, W_{r-1r}),$$

er O.A.S.S. test med nivå ϵ .

Vi kan dermed nytte prosedyren for symmetri (G_2) frå kap. 3.2 med \tilde{Z} erstatta med \tilde{Z}_3 , der

$$\tilde{Z}_3 = \max_x \left(\sum_{i < j} \sum (a_i - a_j) x_{ij} : \sum_{i < j} \sum (a_i - a_j) x_{ij} < Z_3 \right).$$

3.6. Diagonal-parameter symmetri

Vi har diagonal-parameter symmetri i ein rxr-tabell ($r \geq 3$) dersom

$$(3.18) \quad p_{ij}/p_{ji} = \delta_t, \quad t = j-i, \quad i < j.$$

Dette symmetriomgrepet blei presentert og drøfta av Goodman (1979). Parameteren δ_t uttrykkjer odds for at ein observasjon vil falle i ei av cellene (i, j) der $j-i = t$ heller enn i ei av cellene der $j-i = -t$ for $t = 1, 2, \dots, r-1$. Goodman (1979) peikar på at modellar med diagonal-parameter symmetri er spesielt eigna for analyse av rxr-tabellar når kategoriane til kjennemerka har ei naturleg ordning.

Under ein modell med diagonal-parameter symmetri er forholdet mellom talet på overgangar frå i til j og j til i det same for alle (i, j) som er slik at $j-i = t$. Vi skal no vise at modellen for diagonal-parameter symmetri er på formen $\mathcal{P}_{0,B}$ gitt ved (2.5).

La $\xi_t = \log \delta_t$, $t = 1, 2, \dots, r-1$. Ved å nytte (2.3) får vi for (3.18) uttrykt ved dei minimalkanoniske parametrane $\theta_{ij} = \log p_{ij}/p_{rr}$,

$$(3.19) \quad \theta_{ij} - \theta_{ji} = \xi_{j-i}, \quad i < j$$

La B vere gitt ved (3.8) der $h = r^2$ og $d_1 = r(r-1)/2$.

La vidare

$$\xi = (\xi_1, \xi_2, \dots, \xi_{r-1})'$$

dvs. $d_0 = r-1$ (d_0 er definert i kap. 2).

Ved å la $v = 0$ og

$$U = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & \dots & r-1 & r \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ \vdots \\ r-1 \\ r \\ \vdots \\ 2r-3 \\ \vdots \\ d_1 \end{matrix} & \begin{pmatrix} a_{11} & 1 & 0 & \dots & 0 & 0 \\ a_{12} & 0 & 1 & & & 0 \\ \vdots & \vdots & & & & \\ a_{1r-1} & 0 & 0 & & & 0 \\ a_{21} & 1 & 0 & & & 0 \\ \vdots & \vdots & & & & \\ a_{2r-2} & 0 & 0 & & 1 & 0 \\ \vdots & \vdots & & & & \\ a_{r-11} & 1 & 0 & & 0 & 0 \end{pmatrix} \end{matrix}$$

der a_{ij-i} , $i < j$ er faste reelle tal med $a_{1j-1} = 0$, $j = 2, 3, \dots, r$, følger at

$$B\theta = U \binom{0}{\xi}$$

er ekvivalent med (3.19), dvs. modellen for diagonal-parameter symmetri er på formen $\mathcal{P}_{0,B}$ gitt ved (2.5).

Vi skal etablere ein test for hypotesen $\mathcal{P}_{0,B}$ under modellen $\mathcal{P}_{\gamma,B}$ der ξ , B , U og v er gitt som ovafor. Vi har dermed lagt følgjande apriori restriksjon på parameterrommet,

$$B\theta = U \binom{\gamma}{\xi}$$

\Leftrightarrow

$$(3.20) \quad \theta_{ij} - \theta_{ji} = \xi_{j-i} + a_{ij-i}\gamma, \quad i < j$$

der $a_{1j-1} = 0$, $j = 2, 3, \dots, r$

\Leftrightarrow

$$p_{ij}/p_{ji} = \delta_t \exp(a_{it}\gamma), \quad t = j-i, \quad i < j$$

der $a_{1j-1} = 0$, $j = 2, 3, \dots, r$.

$\mathcal{P}_{\gamma,B}$ er altså \mathcal{P} (definert ved (3.6) med $s=r$) avgrensa ved (3.20).

Ved å setje inn for (3.20) i (3.6) får vi at $\mathcal{P}_{\gamma,B}$ er gitt ved

$$(3.21) \quad dP_{\theta} = a(\theta) \exp\left(\gamma \sum_{1 < i < j} a_{ij-i} x_{ij} + \sum_{i=1}^{r-1} \theta_{ii} x_{ii} + \sum_{i < j} \theta_{ji} (x_{ij} + x_{ji})\right)$$

$$+ \sum_{t=1}^{r-1} \xi_t \sum_{i=1}^{r-t} x_{ii+t} \Big) dP_0.$$

Setning 2 gir da at testen som forkastar

$H : \gamma = 0$ (diagonal-parameter symmetri) mot $A : \gamma > 0$ når

$$Z_4 > k(V, W, X_{11}, X_{22}, \dots, X_{r-1r-1})$$

$$\text{der} \quad Z_4 = \sum_{1 < i < j} a_{ij-i} X_{ij} = \sum_{t=1}^{r-2} \sum_{i=2}^{r-t} a_{it} X_{ii+t}, \quad W_{ij} = X_{ij} + X_{ji},$$

$$V_t = \sum_{i=1}^{r-t} X_{ii+t}, \quad W = (W_{12}, \dots, W_{1r}, W_{23}, \dots, W_{2r}, \dots, W_{r-1r})$$

og $V = (V_1, V_2, \dots, V_{r-1})$, er O.A.S.S. test med nivå ε .

Vi skal vise nedafor at k berre er avhengig av V og W . Dermed blir testen som forkastar

$H : \gamma = 0$ mot $A : \gamma > 0$

når

$$Z_4 > k(V, W)$$

O.A.S.S. test med nivå ε .

Vi skal no vise at nullfordelinga til testen er konvolusjonen av $r-2$ multivariate hypergeometriske fordelingar.

La $x = (x_{ij} : 1 < i < j, i = 2, 3, \dots, r-1; j = 3, 4, \dots, r)$ og

$$E = \{x : \sum_{i=1}^{r-t} x_{ii+t} = v_t, t = 1, 2, \dots, r-1, x_{ij} + x_{ji} = w_{ij}, i < j\}.$$

Når H er oppfylt får vi frå (3.21),

$$\begin{aligned} (3.22) \quad & \Pr\left(\bigcap_{t=1}^{r-1} (V_t = v_t), \bigcap_{i < j} (W_{ij} = w_{ij}), \bigcap_{i=1}^{r-1} (X_{ii} = x_{ii})\right) \\ &= \sum_{x \in E} N! \left[\prod_{i=1}^r \prod_{j=1}^r x_{ij}! \right]^{-1} \exp\left(\sum_{i=1}^{r-1} \theta_{ii} x_{ii} + \sum_{i < j} \theta_{ji} w_{ij} + \sum_{t=1}^{r-1} \xi_t v_t\right) \\ &= N! \left[\prod_{i=1}^{r-1} x_{ii}! v_{r-1}! (w_{1r}^{-v_{r-1}})!(N - \sum_{i=1}^{r-1} x_{ii} - \sum_{i < j} w_{ij})! \right]^{-1} \\ &\quad \cdot \left[\exp\left(\sum_{i=1}^{r-1} \theta_{ii} x_{ii} + \sum_{i < j} \theta_{ji} w_{ij} + \sum_{t=1}^{r-1} \xi_t v_t\right) \sum_{\substack{1 < i < j \\ x_{ij}}} \left[\prod_{1 < i < j} x_{ij}! (w_{ij}^{-x_{ij}})! \right] \right. \\ &\quad \left. \cdot \prod_{t=1}^{r-2} (v_t^{-\sum_{i=2}^{r-t} x_{ii+t}})!(w_{1t+1}^{-v_t + \sum_{i=2}^{r-t} x_{ii+t}})! \right]^{-1} \end{aligned}$$

Vidare har vi

$$\begin{aligned} & \sum_{x_{r-tr} \dots x_{r-t-1r-1}} \binom{w_{r-tr}}{x_{r-tr}} \sum_{x_{r-t-1r-1} \dots x_{r-t-1r-1}} \binom{w_{r-t-1r-1}}{x_{r-t-1r-1}} \dots \sum_{x_{33+t} \dots x_{33+t}} \binom{w_{33+t}}{x_{33+t}} \\ & \cdot \sum_{x_{22+t} \dots x_{22+t}} \binom{w_{22+t}}{x_{22+t}} \binom{w_{11+t}}{x_{11+t}} = \binom{\sum_{i=1}^{r-t} w_{ii+t}}{v_t}, t = 1, 2, \dots, r-2. \end{aligned}$$

som gir

$$\begin{aligned} (3.23) \quad & \left(\prod_{1 < i < j} w_{ij}! \prod_{t=1}^{r-2} w_{1t+1} \right) \sum_{1 < i < j} \left[\prod_{1 < i < j} x_{ij}! (w_{ij}^{-x_{ij}})! \right. \\ & \left. \cdot \prod_{t=1}^{r-2} (v_t^{-\sum_{i=2}^{r-t} x_{ii+t}})!(w_{1t+1}^{-v_t + \sum_{i=2}^{r-t} x_{ii+t}})! \right]^{-1} \end{aligned}$$

$$\begin{aligned}
&= \sum_{1 < i < j} \prod_{1 < i < j} (w_{ij})^{r-2} \prod_{t=1}^{w_{1t+1}} \left(v_t - \sum_{i=2}^{r-t} x_{ii+t} \right) \\
&= \left[\begin{array}{cc} w_{2r} & w_{1r-1} \\ \sum_{x_{2r}} (x_{2r}) & v_{r-2} - x_{2r} \end{array} \right] \left[\begin{array}{ccc} w_{3r} & w_{2r-1} & w_{1r-2} \\ \sum_{x_{3r}} (x_{3r}) & \sum_{x_{2r-1}} (x_{2r-1}) & v_{r-3} - x_{2r-1} - x_{3r} \end{array} \right] \\
&\dots \dots \left[\begin{array}{cccc} w_{r-1r} & w_{r-2r-1} & \dots & w_{34} \\ \sum_{x_{r-1r}} (x_{r-1r}) & \sum_{x_{r-2r-1}} (x_{r-2r-1}) & \dots & \sum_{x_{34}} (x_{34}) \end{array} \right] \\
&\left[\begin{array}{cc} w_{23} & w_{12} \\ \sum_{x_{23}} (x_{23}) & v_1 - \sum_{i=2}^{r-1} x_{ii+1} \end{array} \right] \\
&= \left(\begin{array}{ccc} w_{1r-1} + w_{2r} & w_{1r-2} + w_{2r-1} + w_{3r} & \dots + w_{r-1r} \\ v_{r-2} & v_{r-3} & v_1 \end{array} \right) \\
&= \prod_{t=1}^{r-t} \left(v_t - \sum_{i=1}^{r-2} w_{ii+t} \right)
\end{aligned}$$

Ved å nytte (3.22) og (3.23) får vi for den betingede fordelinga til X_{ij} , $1 < i < j$, $i, j = 2, 3, \dots, r$ gitt V, W og X_{ii} , $i = 1, 2, \dots, r-1$

$$\begin{aligned}
(3.24) \quad & \Pr \left(\bigcap_{1 < i < j} (X_{ij} = x_{ij}) \mid \bigcap_{t=1}^{r-1} (V_t = v_t), \bigcap_{i < j} (W_{ij} = w_{ij}), \bigcap_{i=1}^{r-1} (X_{ii} = x_{ii}) \right) \\
&= \frac{\Pr \left(\bigcap_{1 < i < j} (X_{ij} = x_{ij}), \bigcap_{t=1}^{r-1} (V_t = v_t), \bigcap_{i < j} (W_{ij} = w_{ij}), \bigcap_{i=1}^{r-1} (X_{ii} = x_{ii}) \right)}{\Pr \left(\bigcap_{t=1}^{r-1} (V_t = v_t), \bigcap_{i < j} (W_{ij} = w_{ij}), \bigcap_{i=1}^{r-1} (X_{ii} = x_{ii}) \right)} \\
&= \frac{\left[\prod_{1 < i < j} x_{ij}! (w_{ij} - x_{ij})! \prod_{t=1}^{r-2} (v_t - \sum_{i=2}^{r-t} x_{ii+t})! (w_{1t+1} - v_t + \sum_{i=2}^{r-t} x_{ii+t})! \right]^{-1}}{\left[\prod_{1 < i < j} w_{ij}! \prod_{t=1}^{r-2} w_{1t+1} \right]^{-1} \prod_{t=1}^{r-2} \left(\sum_{i=1}^{r-2} w_{ii+t} \right) v_t}
\end{aligned}$$

$$\begin{aligned}
 & \prod_{1 < i < j} \prod_{t=1}^{r-2} \left(\frac{w_{ij}^{r-2}}{x_{ij}^{t-1}} \right) \prod_{t=1}^{r-t} \left(\frac{w_{1t+1}}{v_t - \sum_{i=2}^{r-t} x_{ii+t}} \right) \\
 &= \frac{\prod_{t=1}^{r-2} \prod_{i=1}^{r-2-t} w_{ii+t}}{v_t} \\
 &= \frac{w_{2r}^{r-2} w_{1r-1}^{r-3} \dots w_{12}^2}{v_{r-2} v_{r-3} \dots v_1} \cdot \frac{w_{3r}^{r-3} w_{2r-1}^{r-4} \dots w_{1r-2}^2}{v_{r-3} v_{r-4} \dots v_1} \dots \dots \dots \\
 & \dots \dots \dots \frac{w_{r-1r}^{r-1} w_{r-2r-1}^{r-2} \dots w_{12}^2}{v_1} \cdot \frac{x_{r-1r}^{r-1} x_{r-2r-1}^{r-2} \dots x_{12}^2}{v_1}
 \end{aligned}$$

der $x_{1t+1} = v_t - \sum_{i=2}^{r-t} x_{ii+t}, \quad t = 1, 2, 3, \dots, r-2.$

vi har dermed vist at k er uavhengig av $X_{ii}, i = 1, 2, \dots, r-1.$

I praksis forkaster vi hypotesen om diagonal-parameter symmetri, $\gamma = 0$ mot $\gamma > 0$ dersom

$$G_4(\tilde{Z}_4 | V, W) = \sum_{\ell \leq \tilde{Z}_4} g_4(\ell | V, W) > 1 - \epsilon$$

der $\tilde{Z}_4 = \max_x \left(\sum_{1 < i < j} a_{ij-i} x_{ij} : \sum_{1 < i < j} a_{ij-i} x_{ij} < Z_4 \right),$

$$g_4(z_4 | v, w) = \sum_{x \in D} \prod_{t=1}^{r-2} \left[\frac{v_t - \sum_{i=2}^{r-t} x_{ii+t}}{r-t} \prod_{i=2}^{r-t} \left(\frac{w_{ii+t}}{x_{ii+t}} \right) \right]$$

og

$$D = \{x : \sum_{1 < i < j} a_{ij-i} x_{ij} = z_4\}.$$

Sidan den betingta fordelinga (3.24) er produktet av multivariate hypergeometriske fordelingar, kan vi rekne ut G_4 ved hjelp av program for den multivariate hypergeometriske fordelinga og den diskrete konvolusjonen.

MERK: Når $r = 3$ blir (3.24) den hypergeometriske fordelinga. I denne situasjonen blir derfor metoden i dette avsnittet analog Fisher-Irwin's test for uavhengighet i 2×2 -tabellar.

4. EKSakte metodar for ulike hypotesar i ein to-vegstabell når modellen er produktmultinomisk

I dette kapitlet skal vi sjå på $r \times s$ -tabellen som resultat av r multinomiske forsøksrekker, der verdiane til kjennemerket i forspalten av tabellen er gitte. I denne situasjon får vi tabellen

	B_1	B_2	\dots	B_j	\dots	B_s	Sum
A_1	X_{11}	X_{12}	\dots	X_{1j}	\dots	X_{1s}	N_1
A_2							
\vdots							
\vdots							
A_i	X_{i1}	X_{i2}	\dots	X_{ij}	\dots	X_{is}	N_i
\vdots							
\vdots							
A_r	X_{r1}	X_{r2}	\dots	X_{rj}	\dots	X_{rs}	N_r
Sum	X_{+1}	X_{+2}	\dots	X_{+j}	\dots	X_{+s}	N

der N_1, N_2, \dots, N_r er gitte tal og $X_{i1}, X_{i2}, \dots, X_{is}$ er multinomisk fordelt med parametrar $q_{i1}, q_{i2}, \dots, q_{is}$; $q_{i+} = \sum_{j=1}^s q_{ij} = 1, i = 1, 2, \dots, r$. Vi har derfor at modellen for tabellen er produktmultinomisk.

Forskjellen på denne tabellen og tabellen i kap. 3, er at vi no føreset at marginalsammene $X_{i+} = N_i, i = 1, 2, \dots, r$ er kjende/fastlagde før vi har samla inn datamaterialet. Denne tabellen vil altså vise korleis einingar frå r ulike populasjonar (A_1, \dots, A_r) fordeler seg etter eit kjennemerke med s kategoriar (B_1, \dots, B_s).

4.1. Homogenitet

Vi ønskjer å vurdere tabellen med utgangspunkt i hypotesen om homogenitet;

$$(4.1) \quad q_{1j} = q_{2j} = \dots = q_{rj}, \quad j = 1, 2, \dots, s.$$

(4.1) uttrykkjer at sannsyna for B_1, B_2, \dots, B_s har halde seg konstante gjennom alle forsøksrekkene.

Dette er ei generalisering av den situasjonen som blir omtala i kap. 3.4., Aaberge (1979).

Vi har at

$$(4.2) \quad \frac{q_{ij} \cdot q_{rs}}{q_{rj} \cdot q_{is}} = 1, \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1$$

er ekvivalent med (4.1).

Resultata som blei presenterte i kap. 2, gjeld for tabellar som er resultat av ei multinomisk forsøksrekkje; dvs. begge kjennemerka i tabellen er stokastiske. I dette kapitlet føreset vi at dei r verdiane til kjennemerket i forspalten til tabellen er kjende; tabellen er resultatet av r multinomiske forsøksrekkjer.

Modellen for ein rxs -tabell som er resultatet av r multinomiske forsøksrekkjer, er som kjent gitt ved

$$(4.3) \quad \Pr\left(\bigcap_{i=1}^r \bigcap_{j=1}^s (X_{ij} = x_{ij})\right) = \prod_{i=1}^r \left(\frac{N_i!}{\prod_{j=1}^s x_{ij}!} \prod_{j=1}^s q_{ij}^{x_{ij}} \right)$$

der

$$\sum_{j=1}^s q_{ij} = 1, \quad i = 1, 2, \dots, r.$$

Vi har følgjande samanheng mellom q_{ij} , $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$ og parametrane p_{ij} , $i=1,2,\dots,r$; $j=1,2,\dots,s$ definert i kap. 3,

$$q_{ij} = \frac{p_{ij}}{p_{i+}}, \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, s.$$

Dette gir

$$\frac{q_{ij} \cdot q_{rs}}{q_{is} \cdot q_{rj}} = \frac{p_{ij} \cdot p_{rs}}{p_{is} \cdot p_{rj}} = \alpha_{ij}.$$

Uavhengigheit for tabellen i kap. 3 er derfor ekvivalent med homogenitet for tabellen i dette kapitlet. Vi har dessutan at den produktmultinomiske modellen (4.3) kan uttrykkjast som den betinga multi-

nomiske modellen i kap. 3 gitt marginalsammene $X_{i+} = N_i$, $i = 1, 2, \dots, r$. Vi kan dermed vurdere homogenitet for tabellen i dette kapitlet ved å ta i bruk resultatene vi kom fram til for uavhengighet for tabellen i kap. 3.

I denne situasjonen har vi at parametrane η_{ij} gitt ved

$$\eta_{ij} = \log \frac{q_{ij}}{q_{is}}, \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, s-1$$

er minimalkanoniske.

Uttrykt ved dei minimalkanoniske parametrane (θ_{ij}) for modellen i kap. 3, får vi

$$(4.4) \quad \eta_{ij} = \theta_{ij} - \theta_{is}, \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1$$

$$\eta_{rj} = \theta_{rj}, \quad j = 1, 2, \dots, s-1.$$

Ved å settje inn for (4.4) i (3.5) får vi at homogenitet kan studerast ved hjelp av parameteren γ under restriksjonen

$$(4.5) \quad \eta_{ij} - \eta_{rj} = a_{ij}\gamma, \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1$$

\Leftrightarrow

$$\frac{q_{ij} \cdot q_{rs}}{q_{is} \cdot q_{rj}} = \exp(a_{ij}\gamma), \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1.$$

Vi skal no finne fordelingsklassen (4.3) avgrensa ved (4.5), $\int_{\eta, B}$. Den tilsvarande fordelingsklassen (under studiet av uavhengighet) i kap. 3.1. er gitt ved (3.7). Ved å nytte Lemma 8 (kap. 2) Lehmann (1959) finn vi derfor fordelingsklassen $\int_{\eta, B}$ frå (3.7) ved å betinga m.o.p. marginalsammene $X_{i+} = N_i$, $i = 1, 2, \dots, r$. Dette gir

$$(4.6) \quad dP_{\eta} = a'(\eta) \exp\left(\gamma \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} + \sum_{j=1}^{s-1} \eta_{rj} x_{rj}\right) dP'_{\eta}$$

der P'_{η} er sannsynfordelinga svarande til $q_{i1} = q_{i2} = \dots = q_{is} = \frac{1}{s}$, $i = 1, 2, \dots, r$.

Setning 2 gir derfor at testen som forkastar hypotesen om homogenitet, $\gamma = 0$ (dvs. dei r populasjonane (forsøksrekkjene) har samme fordeling på kategoriane til kjennemerket B) mot alternativa $\gamma > 0$ når

$$Y = \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} X_{ij} > k(X_{+1}, X_{+2}, \dots, X_{+s-1})$$

er O.A.S.S. test med nivå ϵ .

For å fastsetje k , treng vi den betingda fordelinga til Y gitt X_{+1}, \dots, X_{+s-1} for $\gamma = 0$. Frå dei grunngevingane vi har gitt ovafor, følgjer opplagt at

$$\begin{aligned} g_5(y | x_{+1}, x_{+2}, \dots, x_{+s-1}) &= \Pr_{\gamma=0}(Y=y | \bigcap_{j=1}^{s-1} (X_{+j} = x_{+j}), \bigcap_{i=1}^{r-1} (X_{i+} = N_i)) \\ &= g_1(y | N_1, N_2, \dots, N_{r-1}, x_{+1}, x_{+2}, \dots, x_{+s-1}) \\ &= \sum_{x \in S_1} \frac{\binom{N_1}{x_{11}, \dots, x_{1s}} \dots \binom{N_r}{x_{r1}, \dots, x_{rs}}}{\binom{N}{x_{+1}, \dots, x_{+s}}}, \end{aligned}$$

der

$$S_1 = \{x : \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} = y\},$$

$$x_{is} = N_i - \sum_{j=1}^{s-1} x_{ij}, \quad i = 1, 2, \dots, r$$

$$x_{rj} = x_{+j} - \sum_{i=1}^{r-1} x_{ij}, \quad j = 1, 2, \dots, s$$

og

$$\sum_{j=1}^s x_{+j} = \sum_{i=1}^s N_i = N.$$

I praksis vil vi ikkje føreta randomisering. Vi forkastar derfor hypotesen om homogenitet $\gamma = 0$ mot $\gamma > 0$ dersom dei observerte X_{ij} , $i = 1, 2, \dots, r-1$; $j = 1, 2, \dots, s-1$ og X_{+j} , $j = 1, 2, \dots, s-1$ er slik at

$$G_5(\tilde{Y}|X_{+1}, \dots, X_{+s-1}) = \sum_{\ell \leq \tilde{Y}} g_5(\ell|X_{+1}, \dots, X_{+s-1}) > 1-\epsilon.$$

der

$$\tilde{Y} = \max_x \left(\sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} : \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} < Y \right).$$

Testprosedyren for homogenitet er identisk med den vi fann for uavhengigheit i kap. 3.1. Når det gjeld eigenskapar ved metoden, viser vi til dei kommentarar som er gitt under Merk i kap. 3.1.

4.2. Trend

Dersom $s = 2$, blir tabellen på side 39 resultatet av r binomiske forsøksrekkjer. Kjennermerket B har to kategoriar (B_1, B_2). Dersom forspaltevariabelen A har ei naturleg ordning, kan det vere aktuelt å vurdere trend i førekomsten av B_1 . Vi kan uttrykkje trend i modellen (4.5) ved å la a_{ij} , $j = 1, 2, \dots, s-1$ vere veksande eller avtakande funksjonar av i . Når $s = 2$ blir $a_{ij} = a_i$ og $q_{i1} = q_i$ ($q_{i2} = 1 - q_i$) slik at modellen (4.5) uttrykt ved dei binomiske parametrane (q_i) blir

$$(4.7) \quad \frac{q_i(1-q_r)}{q_r(1-q_i)} = \exp(a_i \gamma), \quad i = 1, 2, \dots, r-1.$$

Ved å la $a_i = -1/i$ eller $a_i = 1/i$ får vi vurdert om det er respektive veksande eller avtakande tendens i tabellen ved å teste hypotesen $\gamma = 0$ mot alternativa $\gamma > 0$. Hypotesen $\gamma = 0$ uttrykkjer at førekomsten av B_1 relativt til B_2 er den same i dei r populasjonane. $a_i = -1/i$ og $\gamma > 0$ medfører

$$(i) \quad q_1 < q_2 < \dots < q_r$$

medan $a_i = 1/i$ og $\gamma > 0$ medfører

$$(ii) \quad q_1 > q_2 > \dots > q_r.$$

Dette følgjer opplagt frå (4.7).

For å vurdere desse trendane kan vi nytte den eksakte metoden ovafor; dvs. vi kan påstå (i) eller (ii) med a_i lik respektive $-1/i$ og $1/i$ når

$$\sum_{\ell \in Y} g_5(\ell | X_{+1}) > 1 - \epsilon$$

der

$$g_5(y | x_{+1}) = \sum_{x \in S_1} \frac{\binom{N_1}{x_{11}} \dots \binom{N_r}{x_{r1}}}{\binom{N}{x_{+1}}},$$

$$x = (x_{11}, x_{21}, \dots, x_{r-11}), S_1 = \{x : \sum_{i=1}^{r-1} a_i x_{i1} = y\},$$

$$x_{r1} = x_{+1} - \sum_{i=1}^{r-1} x_{i1} \quad \text{og} \quad \sum_{i=1}^r N_i = N.$$

I tabellar med s større enn 2, kan det vere aktuelt å studere trend i cellesannsyna etter kategoriane B_j , $j = 1, 2, \dots, s-1$. Vi føreset at kategoriane til kjennemerket i forspalten har ei naturleg ordning. Denne type trendhypotesar kan vi få uttrykt i modellen (4.5) ved å la $a_{ij} = a_i$, $j = 1, 2, \dots, s-1$, der a_i er ein veksande negativ funksjon av i eller avtakande positiv funksjon av i . Vi skal no grunngi dette formelt.

La

$$a_{ij} = \begin{cases} a_i, & j = 1, 2, \dots, s-1 \\ 0 & j = s \end{cases} \quad i = 1, 2, \dots, r-1$$

Frå (4.5) får vi

$$(4.7) \quad q_{ij} q_{rs} = q_{is} q_{rj} \exp(a_{ij} \gamma), \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s.$$

Sidan

$$\sum_{j=1}^s q_{ij} = 1, \quad i = 1, 2, \dots, r, \quad \text{får vi ved å summere over } j \text{ på begge}$$

sider i (4.7)

$$q_{rs} = q_{is} \sum_{j=1}^s q_{rj} \exp(a_{ij} \gamma), \quad i = 1, 2, \dots, r-1.$$

Innsett i (4.7) gir dette

$$(4.8) \quad \frac{q_{ij}}{q_{rj}} = \frac{\exp(a_i \gamma)}{\sum_{j=1}^s q_{rj} \exp(a_{ij} \gamma)} = \frac{\exp(a_i \gamma)}{q_{rs} + (1 - q_{rs}) \exp(a_i \gamma)}$$

$$= \frac{1}{1 - q_{rs} + q_{rs} \exp(-a_i \gamma)}, \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1.$$

Sidan $\sum_{j=1}^s q_{rj} = 1$, følger frå (4.8) at " $\gamma=0$ " er ekvivalent med

$$q_{1j} = q_{2j} = \dots = q_{rj}, \quad j = 1, 2, \dots, s.$$

For $\gamma > 0$ viser (4.8) at q_{ij} er ein veksande funksjon i a_i og at $q_{ij} < q_{rj}$ (for alle i) viss og berre viss a_i er ein negativ funksjon av i .
Dermed har vi vist at modellen (4.5) med $\gamma > 0$ og $a_{ij} = a_i \in \mathcal{A}_1$ der

$$\mathcal{A}_1 = \{a(i) : a(i) < 0 \text{ og } a(i) \text{ monotont veksande funksjon av } i\}$$

medfører

$$q_{1j} < q_{2j} < \dots < q_{rj}, \quad j = 1, 2, \dots, s-1.$$

Ved å la $a_{ij} = a_i \in \mathcal{A}_2$ der

$$\mathcal{A}_2 = \{a(i) : a_i > 0 \text{ og } a(i) \text{ monotont avtakande funksjon av } i\}$$

finn vi på tilsvarande måte at

$$q_{1j} > q_{2j} > \dots > q_{rj}, \quad j = 1, 2, \dots, s-1.$$

For å vurdere problem av typen

$$H : q_{1j} = q_{2j} = \dots = q_{rj}, j = 1, 2, \dots, s-1$$

mot

$$A_1 : q_{1j} < q_{2j} < \dots < q_{rj}, j = 1, 2, \dots, s-1$$

eller

$$A_2 : q_{1j} > q_{2j} > \dots > q_{rj}, j = 1, 2, \dots, s-1$$

kan vi derfor nytte metoden frå kap. 4.1. med $a_{ij} = a_i$ gitt ved funksjonar frå respektive \mathcal{A}_1 og \mathcal{A}_2 .

I enkelte situasjonar vil det vere naturleg å studere trend i oddsforholda α_{ij} , der

$$\alpha_{ij} = \frac{q_{ij} \cdot q_{rs}}{q_{is} \cdot q_{rj}} = \frac{q_{ij}/q_{is}}{q_{rj}/q_{rs}}, i = 1, 2, \dots, r-1; j = 1, 2, \dots, s-1.$$

Når $\alpha_{ij} > 1$ har vi at førekomsten av B_j relativt B_s er større for populasjonen A_i enn for populasjonen A_r .

Dersom tabellen omtalte sysselsetjingstatusen til ulike aldersgrupper, kunne f.eks. α_{ij} gi uttrykk for forholdet mellom førekomsten av "sysselsetjing" relativt "ikkje-sysselsetjing" i to aldersgrupper.

La \mathcal{A}_1 og \mathcal{A}_2 vere definert som ovafor. Ved å la

$$\alpha_{ij} = \exp(a_{ij}\gamma), \gamma > 0$$

der

$$a_{ij} \in \mathcal{A}_1, j = 1, 2, \dots, b$$

og

$$a_{ij} \in \mathcal{A}_2, j = b+1, b+2, \dots, s-1,$$

blir

$$\alpha_{1j} < \alpha_{2j} < \dots < \alpha_{rj}, j = 1, 2, \dots, b$$

og

$$\alpha_{1j} > \alpha_{2j} > \dots > \alpha_{rj}, j = b+1, b+2, \dots, s-1.$$

Dermed følgjer at metoden i kap. 4.1 kan nyttast til å vurdere trend i oddsforholda, når forspalten i tabellen har ei naturleg ordning.

4.3. Ulike problem i 2xs-tabellen

Vi skal gi ei spesiell handsaming av rxs-tabellen ovafor for $r = 2$. Modellen for tabellen er da gitt ved produktet av to multinomiske fordelingar. I tilknytning til denne tabellen skal vi drøfte:

- (I) problem uttrykt ved alle parametrane i modellen
- (II) problem uttrykt ved ei delmengd av parameterrommet.

(I)

Vi skal først ta føre oss problem av typen

$$H : q_{1j} = q_{2j}, j = 1, 2, \dots, s$$

mot

$$A : q_{1j} \begin{cases} < q_{2j}, j = 1, 2, \dots, t \\ > q_{2j}, j = t+1, t+2, \dots, s-1. \end{cases}$$

Alternativet (A) gir uttrykk for at t av cellene i forsøksrekke 2 har større sannsynsmasse enn dei tilsvarende cellene i forsøksrekke 1, medan i $s-t-1$ av cellene er sannsynsmassen størst i forsøksrekke 1.

Vi skal vise at problem av denne typen kan uttrykkjast i modellen (4.5).

Ved å la $a_{1j} = a_j$ i (4.5), får vi

$$(4.9) \quad q_{1j} \cdot q_{2s} = q_{2j} \cdot q_{1s} \exp(a_j \gamma), j = 1, 2, \dots, s-1.$$

Det følgjer frå det vi tidlegare har vist at " $\gamma=0$ " er ekvivalent med

$$q_{1j} = q_{2j}, j = 1, 2, \dots, s.$$

La $\gamma > 0$ og

$$a_j = \begin{cases} b_1 < 0, j = 1, 2, \dots, t \\ b_2 > 0, j = t+1, t+2, \dots, s-1. \end{cases}$$

Da gir (4.9)

$$(4.10) \quad q_{1j}/q_{2j} \begin{cases} < q_{1s}/q_{2s}, j = 1, 2, \dots, t \\ > q_{1s}/q_{2s}, j = t+1, t+2, \dots, s-1. \end{cases}$$

og

(4.11) $q_{1j}q_{2\ell} = q_{2j}q_{1\ell}$ for alle kombinasjoner av j og ℓ når $j, \ell = 1, 2, \dots, t$
 og $j, \ell = t+1, t+2, \dots, s-1$.

Frå (4.10) får vi vidare

$$q_{1j}(1-q_{21}-q_{22}-\dots-q_{2s-1}) < q_{2j}(1-q_{11}-q_{12}-\dots-q_{1s-1}), \quad j = 1, 2, \dots, t$$

som igjen er ekvivalent med

$$\begin{aligned} q_{1j} - q_{2j} &< q_{1j} \sum_{\ell=1}^{s-1} q_{2\ell} - q_{2j} \sum_{\ell=1}^{s-1} q_{1\ell} \\ &= \sum_{\ell=1}^t (q_{1j}q_{2\ell} - q_{2j}q_{1\ell}) + \sum_{\ell=t+1}^{s-1} (q_{1j}q_{2\ell} - q_{2j}q_{1\ell}), \quad j = 1, 2, \dots, t. \end{aligned}$$

Frå (4.11) følgjer at

$$\sum_{\ell=1}^t (q_{1j}q_{2\ell} - q_{2j}q_{1\ell}) = 0, \quad j = 1, 2, \dots, t \text{ medan (4.10) gir at}$$

$$\sum_{\ell=t+1}^{s-1} (q_{1j}q_{2\ell} - q_{2j}q_{1\ell}) = \sum_{\ell=t+1}^{s-1} q_{2j}q_{2\ell} (q_{1j}/q_{2j} - q_{1\ell}/q_{2\ell}) < 0,$$

$$j = 1, 2, \dots, t.$$

Altså blir

$$q_{1j} - q_{2j} < 0, \quad j = 1, 2, \dots, t.$$

Frå (4.9) følgjer også at

$$q_{1j} - q_{2j} > \sum_{\ell=1}^t (q_{1j}q_{2\ell} - q_{2j}q_{1\ell}) + \sum_{\ell=t+1}^{s-1} (q_{1j}q_{2\ell} - q_{2j}q_{1\ell}),$$

$$j = t+1, t+2, \dots, s-1.$$

(4.11) gir at

$$\sum_{\ell=t+1}^{s-1} (q_{1j}q_{2\ell} - q_{2j}q_{1\ell}) = 0, \quad j = t+1, t+2, \dots, s-1$$

medan (4.10) gir at

$$\sum_{\ell=1}^t (q_{1j}q_{2\ell} - q_{2j}q_{1\ell}) > 0, \quad j = t+1, t+2, \dots, s-1.$$

Altså blir

$$q_{1j} - q_{2j} > 0, \quad j = t+1, t+2, \dots, s-1.$$

Vi har dermed vist at $A' \subset A$, der A' er mengda av q gitt ved (4.9) med $\gamma > 0$ og

$$a_j = \begin{cases} b_1 < 0, & j = 1, 2, \dots, t \\ b_2 > 0, & j = t+1, t+2, \dots, s-1, \end{cases}$$

der $q = (q_{11}, q_{12}, \dots, q_{1s}, q_{21}, q_{22}, \dots, q_{2s})'$

Da følgjer frå resultatata i kap. 4.1 at testen som forkastar H mot A' når

$$Y = \sum_{j=1}^{s-1} a_j X_{1j} > k(X_{+1}, \dots, X_{+s-1}) \text{ er O.A.S.S. test med nivå } \varepsilon.$$

Ved bruk av metoden nyttar vi den prosedyren (G_5) som blei omtala i kap. 4.1.

Tre-desisjonsutgåva (sjå kap. 2.1) av metoden vert derfor;
påstå $\gamma > 0$

når

$$G_5(\tilde{Y} | X_{+1}, \dots, X_{+s-1}) > 1 - \varepsilon$$

og

påstå $\gamma < 0$

når

$$G_5(Y | X_{+1}, \dots, X_{+s-1}) < \varepsilon.$$

$$\tilde{Y} = \max_x \left(\sum_{j=1}^{s-1} a_j x_{1j} : \sum_{j=1}^{s-1} a_j x_{1j} < Y \right).$$

Den betingta fordelinga til Y gitt X_{+1}, \dots, X_{+s-1} under H blir i denne situasjonen gitt ved

$$g_5(y|x_{+1}, \dots, x_{+s-1}) = \sum_{x \in S_1} \frac{\binom{N_1}{x_{11}, \dots, x_{1s}} \binom{N_2}{x_{21}, \dots, x_{2s}}}{\binom{N}{x_{+1}, \dots, x_{+s}}}$$

$$= \sum_{x \in S_1} \frac{\frac{N_1!}{\prod_{j=1}^s x_{1j}!} \frac{N_2!}{\prod_{j=1}^s x_{2j}!}}{\frac{N!}{\prod_{j=1}^s x_{+j}!}} = \sum_{x \in S_1} \frac{\prod_{j=1}^s \frac{x_{+j}!}{x_{1j}! (x_{+j} - x_{1j})!}}{\frac{N!}{N_1! (N - N_1)!}}$$

$$= \sum_{x \in S_1} \frac{\prod_{j=1}^s \binom{x_{+j}}{x_{1j}}}{\binom{N}{N_1}}$$

der $x_{+s} = N - \sum_{j=1}^{s-1} x_{+j}$, $x_{1s} = N_1 - \sum_{j=1}^{s-1} x_{1j}$ og $S_1 = \{x : \sum_{j=1}^{s-1} a_j x_{1j} = y\}$.

$\left[\binom{N}{N_1} \right]^{-1} \prod_{j=1}^s \binom{x_{+j}}{x_{1j}}$ er den multivariate hypergeometriske fordelinga.

Eksempel 1. I kap. 1.1. presenterte vi analysen av tabell 1 som eit tre-desisjonsproblem. Nedafor skal vi omtale korleis dette kan bli uttrykt ved parametrane i den produktmultinomiske modellen for tabellen.

Observasjonane for dagpendlarane og veke- og langtidspendlarane er kvar for seg multinomisk fordelte med parametrar

	Skiftarbeid	Fast nattarbeid	Dagarbeid
Veke- og langtidspendlarar	q_{11}	q_{12}	q_{13}
Dagpendlarar	q_{21}	q_{22}	q_{23}

der $\sum_{j=1}^3 q_{ij} = 1$, $i = 1, 2$.

La

$$\frac{q_{1j}q_{23}}{q_{2j}q_{13}} = \exp(a_j\gamma), \quad j = 1, 2$$

der $a_1 = -a_2 = 1$.

Da følgjer frå det vi har vist i dette avsnittet at

$$\gamma > 0$$

medfører

$$q_{11} > q_{21} \quad \text{og} \quad q_{12} < q_{22}, \quad \text{dvs. avgjerd (i) i kap. 1.1.,}$$

og dessutan at

$$\gamma < 0$$

medfører

$$q_{11} < q_{21} \quad \text{og} \quad q_{12} > q_{22}, \quad \text{dvs. avgjerd (ii) i kap. 1.1.}$$

Vi har dermed redusert problemet til å avgjere om $\gamma > 0$, $\gamma < 0$ eller eventuelt ikkje seie noko. Vi kan derfor nytte tre-desisjonsutgåva av metoden som blei presentert ovafor. Vi krev nivået $\epsilon = 0.05$ og finn

$$G_5(12|X_{+1}, X_{+2}) = 0.9604 > 0.9500 = 1 - \epsilon.$$

Vi kan derfor påstå $\gamma > 0$, dvs. avgjerd (i) i kap. 1.1.

(II)

Under handsaminga av 2xs-tabellen ovafor inngikk alle parametrane frå dei to forsøksrekkjene i problemet. Vi vil no studere problem som kan bli uttrykte ved ei delmengd av parameterrommet; dvs. vårt problem er berre knytta til ein del av dei s kategoriane i hovudet av tabellen.

F.eks. kan problemet vere å avgjere om

$$q_{1j} < q_{2j}, \quad j = 1, 2, \dots, t_1$$

og

$$q_{1j} > q_{2j}, \quad j = t_1+1, t_1+2, \dots, t_2$$

der

$$t_1 < t_2 < s-1.$$

For å vurdere slike problem, kan vi nytte ein testmetode for

$$H_1 : q_{1j} = q_{2j}, \quad j = 1, 2, \dots, t_2$$

mot

$$A_1 : \begin{cases} < q_{2j}, & j = 1, 2, \dots, t_1 \\ q_{1j} > q_{2j}, & j = t_1+1, t_1+2, \dots, t_2. \end{cases}$$

Det er dermed $2t_2$ parametrar som inngår i problemet.

Som vi peika på i kap. 4.1. er parametrane η_{ij} gitt ved

$$(4.12) \quad \eta_{ij} = \log \frac{q_{ij}}{q_{is}}, \quad i = 1, 2; \quad j = 1, 2, \dots, s-1$$

minimalkanoniske. Vi skal no vise at hypotesen H_1 ikkje kan skrivast på formen

$$B_1 \eta = v_1 + U_1 \begin{pmatrix} 0 \\ \xi \end{pmatrix}$$

der $\eta = (\eta_{11}, \dots, \eta_{1s-1}, \eta_{21}, \dots, \eta_{2s-1})'$, ξ vektor med dimensjon d_0 , U_1 er ei $d_1 \times (d_0+1)$ -matrise med rang d_0+1 ($d_0=0, 1, \dots, d_1-1$), B_1 er ei $d_1 \times (2s-2)$ -matrise med rang d_1 ($d_1=1, 2, \dots, 2s-s$) og v_1 vektor med dimensjon d_1 .

Sidan $\sum_{j=1}^s q_{ij} = 1$, $i = 1, 2$ får vi frå (4.12),

$$q_{ij} = \frac{\exp(\eta_{ij})}{s-1 + \sum_{j=1} \exp(\eta_{ij})}, \quad i = 1, 2; \quad j = 1, 2, \dots, s-1$$

og

$$q_{is} = \frac{1}{s-1 + \sum_{j=1} \exp(\eta_{ij})}, \quad i = 1, 2.$$

Transformasjonen (4.12) er ein-eintydig. Dermed har vi at

H_1

\Leftrightarrow

$$\frac{\exp(\eta_{1j})}{s-1 + \sum_{j=1} \exp(\eta_{1j})} = \frac{\exp(\eta_{2j})}{s-1 + \sum_{j=1} \exp(\eta_{2j})}, \quad j = 1, 2, \dots, t_2$$

\Leftrightarrow

$$\exp(\eta_{1j} - \eta_{2j}) \frac{\sum_{j=1}^{s-1} \exp(\eta_{2j})}{s-1} = 1, \quad j = 1, 2, \dots, t_2$$

\Leftrightarrow

$$\eta_{1j} - \eta_{2j} + \log(1 + \sum_{j=1}^{s-1} \exp(\eta_{2j})) - \log(1 + \sum_{j=1}^{s-1} \exp(\eta_{1j})) = 0,$$

$$j = 1, 2, \dots, t_2.$$

Vi har dermed vist at hypotesen H_1 ikkje er på formen

$$B_1 \eta = v_1 + U_1 \begin{pmatrix} 0 \\ \xi \end{pmatrix}$$

dvs. modellen under hypotesen er ikkje affin og i følgje Lemma 1 (kap. 2) dermed ikkje regulær. Dette hindrar oss i å bruke den same framgangs- måten som tidlegare for å etablere ein testmetode for problemet. Men ved å studere ein affin hypotese som impliserer H_1 , kan vi iallfall nytte resultatata i kap. 2. til å konstruere ein O.A.S.S. test for eit problem generert av den affine delhypotesen til H_1 . Vi vel derfor å ta utgangspunkt i hypotesen

$$H_2 : q_{1j} = q_{2j}, \quad j = 1, 2, \dots, s-1$$

som opplagt impliserer H_1 . Vi skal no konstruere ein test med nivå ε for H_2 mot

$$A_2 : q_{1j} \begin{cases} < q_{2j}, & j = 1, 2, \dots, t_1 \\ > q_{2j}, & j = t_1+1, t_1+2, \dots, t_2 \\ = q_{2j} q_{1s}/q_{2s}, & j = t_2+1, t_2+2, \dots, s-1 \end{cases}$$

og vidare vise at testen har nivå ε mot H_1 ved sida av andre nyttige eigen- skapar.

$$(4.13) \quad \text{La } a_j = \begin{cases} b_1 < 0, & j = 1, 2, \dots, t_1 \\ b_2 > 0, & j = t_1+1, t_1+2, \dots, t_2 \\ 0 & j = t_2+1, t_2+2, \dots, s-1 \end{cases}$$

og la (4.9) vere oppfylt med a_j -ane fastsett ved (4.13). Da følgjer det frå tidlegare at " $\gamma=0$ " og H_2 er ekvivalente utsegn. Vidare følgjer opplagt at

$$H_2 = H_1 \cap \Omega_2 = H_1 \cap \Omega_3$$

der

$$\Omega_2 = \{q : q_{1j} < q_{2j}, j = 1, \dots, t_1; q_{1j} > q_{2j}, j = t_1+1, \dots, t_2; \\ q_{1j} q_{2s} = q_{2j} q_{1s}, j = t_2+1, \dots, s-1 \text{ eller } q_{1j} = q_{2j}, \\ j = 1, \dots, s-1; \sum_{j=1}^s q_{ij} = 1, i = 1, 2 \},$$

$$\Omega_3 = \{q : (4.9) \text{ oppfylt med } a_j\text{-ane gitt ved (4.13)}, \\ \sum_{j=1}^s q_{ij} = 1, i = 1, 2 \text{ og } \gamma \geq 0\}$$

og $q = (q_{11}, q_{12}, \dots, q_{1s}, q_{21}, q_{22}, \dots, q_{2s})'$.

Ved å uttrykkje alternativmengda A_2 ved Ω_2 og H_2 får vi

$$A_2 = \Omega_2 - H_2.$$

La

$$A_3 = \{q : (4.9) \text{ oppfylt med } a_j\text{-ane gitt ved (4.13)}, \\ \sum_{j=1}^s q_{ij} = 1, i = 1, 2 \text{ og } \gamma > 0\},$$

dvs.

$$A_3 = \Omega_3 - H_2.$$

Vi skal no vise at

$$\Omega_3 \subset \Omega_2$$

dvs.

$$A_3 \subset A_2,$$

og deretter gi ein illustrasjon av korleis dei ulike mengdene er plassert i forhold til kvarandre.

La $q \in A_3$. Da følgjer frå (4.9) og (4.13)

$$(4.14) \quad q_{1j}/q_{2j} \begin{cases} < q_{1s}/q_{2s}, & j = 1, 2, \dots, t_1 \\ > q_{1s}/q_{2s}, & j = t_1+1, \dots, t_2 \\ = q_{1s}/q_{2s}, & j = t_2+1, \dots, s-1 \end{cases}$$

og

$$(4.15) \quad q_{1j}q_{2\ell} = q_{2j}q_{1\ell} \text{ for alle kombinasjonar av } j \text{ og } \ell \text{ når}$$

$$j, \ell = 1, 2, \dots, t_1 \text{ og } j, \ell = t_1+1, \dots, t_2.$$

Sidan $q_{is} = 1 - \sum_{\ell=1}^{s-1} q_{i\ell}$, $i = 1, 2$ har vi frå (4.14)

$$q_{1j}(1 - \sum_{\ell=1}^{s-1} q_{2\ell}) < q_{2j}(1 - \sum_{\ell=1}^{s-1} q_{1\ell}), j = 1, 2, \dots, t_1$$

som er ekvivalent med

$$\begin{aligned}
q_{1j} - q_{2j} &< q_{1j} \sum_{\ell=1}^{s-1} q_{2\ell} - q_{2j} \sum_{\ell=1}^{s-1} q_{1\ell} \\
&= \sum_{\ell=1}^{t_1} (q_{1j} q_{2\ell} - q_{2j} q_{1\ell}) + \sum_{\ell=t_1+1}^{t_2} (q_{1j} q_{2\ell} - q_{2j} q_{1\ell}) \\
&+ \sum_{\ell=t_2+1}^{s-1} (q_{1j} q_{2\ell} - q_{2j} q_{1\ell}), \quad j = 1, 2, \dots, t_1.
\end{aligned}$$

(4.15) gir at

$$\sum_{\ell=1}^{t_1} (q_{1j} q_{2\ell} - q_{2j} q_{1\ell}) = 0, \quad j = 1, 2, \dots, t_1$$

medan (4.14) gir at

$$\begin{aligned}
\sum_{\ell=t_1+1}^{t_2} (q_{1j} q_{2\ell} - q_{2j} q_{1\ell}) &= \sum_{\ell=t_1+1}^{t_2} q_{2j} q_{2\ell} (q_{1j}/q_{2j} - q_{1\ell}/q_{2\ell}) < 0, \\
j &= 1, 2, \dots, t_1
\end{aligned}$$

og

$$\begin{aligned}
\sum_{\ell=t_2+1}^{s-1} (q_{1j} q_{2\ell} - q_{2j} q_{1\ell}) &= \sum_{\ell=t_2+1}^{s-1} q_{2\ell} (q_{1j} - q_{2j} q_{1\ell}/q_{2\ell}) \\
&= \sum_{\ell=t_2+1}^{s-1} q_{2\ell} (q_{1j} - q_{2j} q_{1s}/q_{2s}) < 0, \quad j = 1, 2, \dots, t_1.
\end{aligned}$$

Av dette har vi

$$q_{1j} < q_{2j}, \quad j = 1, 2, \dots, t_1.$$

På tilsvarende måte får vi at

$$q_{1j} > q_{2j}, \quad j = t_1+1, \dots, t_2.$$

Vi har altså vist at

$$\underline{q \in A_3 \Rightarrow q \in A_2, \text{ dvs. } A_3 \subset A_2.}$$

La

$$\Omega_1 = \{q : q_{1j} < q_{2j}, \quad j = 1, \dots, t_1;$$

$$q_{1j} > q_{2j}, \quad j = t_1+1, \dots, t_2 \text{ eller } q_{1j} = q_{2j}, \quad j = 1, \dots, t_2;$$

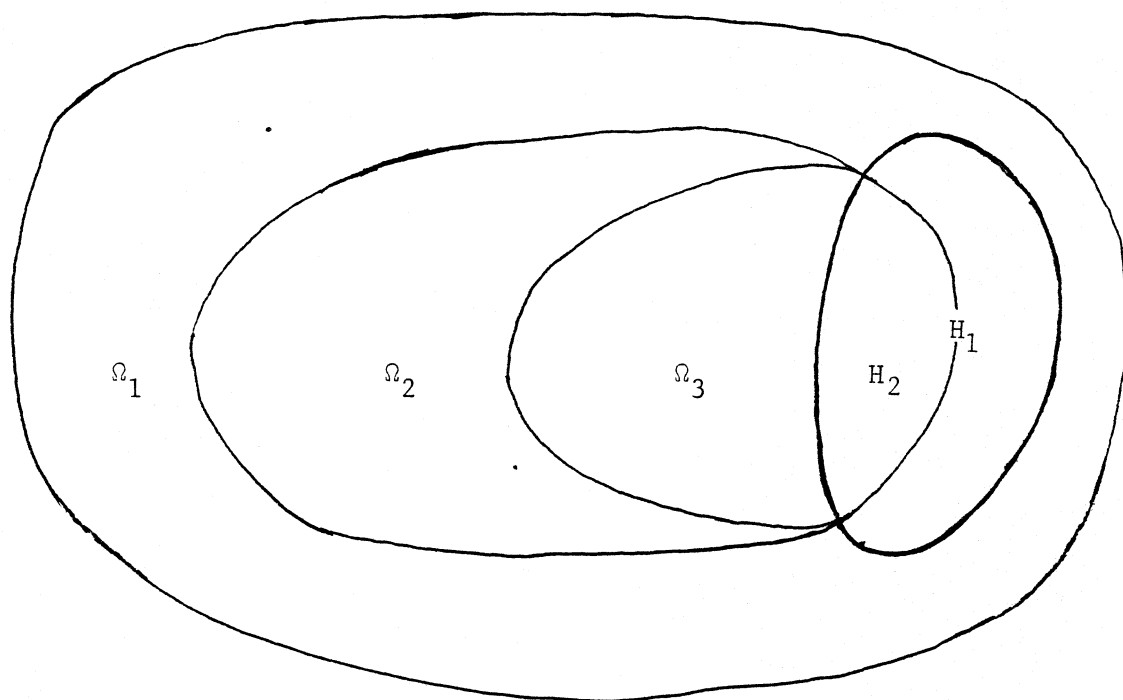
$$\left. \begin{aligned} & \sum_{j=1}^s q_{ij} = 1, \quad i = 1, 2 \end{aligned} \right\}$$

Da blir

$$A_1 = \Omega_1 - H_1.$$

Sidan A_1 opplagt inneheld A_2 , har vi $A_3 \subset A_2 \subset A_1$.

Figur. Illustrasjon av forholda mellom alternativa A_1 , A_2 og A_3 og hypotesane H_1 og H_2 .



Testen som forkastar H_2 mot A_2 med nivå ϵ når

$$Y = \sum_{j=1}^{t_2} a_j X_{1j} > k(X_{+1}, \dots, X_{+t_2})$$

har følgjande nyttige eigenskapar:

- (i) nivå ϵ mot hypotesen H_1
- (ii) overalt sterkast styrkerett mot A_3 .

Prov. Sidan (4.9) (der a_j -ane er gitt ved (4.13)) er med i modellklassen (4.5) for $r = 2$, følger det frå kap. 4.1. at testen som forkastar

$H_2 (\gamma=0)$ mot A_3 når

$$Y = \sum_{j=1}^{t_2} a_j X_{1j} > k(X_{+1}, \dots, X_{+s-1})$$

er O.A.S.S. test med nivå ϵ .

Den betingta fordelinga for Y gitt $(X_{+1}, \dots, X_{+s-1})$ under H_2 er gitt ved

$$g_5(y | x_{+1}, \dots, x_{+s-1}) = \sum_{x \in S_1} \frac{\binom{N_1}{x_{11}, \dots, x_{1s}} \binom{N_2}{x_{21}, \dots, x_{2s}}}{\binom{N}{x_{+1}, \dots, x_{+s}}}$$

$$= \sum_{x \in S_1} \frac{\prod_{j=1}^s \binom{x_{+j}}{x_{1j}}}{\binom{N}{N_1}}$$

Sidan $S_1 = \{x : \sum_{j=1}^{t_2} a_j x_{1j} = y\} = \{x_1 : \sum_{j=1}^{t_2} a_j x_{1j} = y\} \stackrel{\text{def.}}{=} S_1'$

der $x = (x_{11}, \dots, x_{1s})'$ og $x_1 = (x_{11}, \dots, x_{1t_2})'$, får vi vidare

$$g_5(y | x_{+1}, \dots, x_{+s-1}) = \sum_{x_1 \in S_1'} \frac{\prod_{j=1}^s \binom{x_{+j}}{x_{1j}}}{\binom{N}{N_1}} \sum_{x_{1t_2+1}, \dots, x_{1s}} \prod_{j=t_2+1}^s \binom{x_{+j}}{x_{1j}}$$

Vi har at

$$\sum_{x_{1t_2+1}, \dots, x_{1s}} \prod_{j=t_2+1}^s \binom{x_{+j}}{x_{1j}} = \binom{\sum_{j=t_2+1}^s x_{+j}}{s} = \binom{N - \sum_{j=1}^{t_2} x_{+j}}{t_2}$$

$$\sum_{j=t_2+1}^s \binom{x_{+j}}{x_{1j}} = \binom{N_1 - \sum_{j=1}^{t_2} x_{1j}}{t_2}$$

som medfører at

$$g_5(y|x_{+1}, \dots, x_{+s-1}) = \sum_{x_1 \in S_1'} \frac{\binom{x_{+1}}{x_{11}} \dots \binom{x_{+t_2}}{x_{1t_2}} \binom{N - \sum_{j=1}^{t_2} x_{+j}}{N_1 - \sum_{j=1}^{t_2} x_{1j}}}{\binom{N}{N_1}}$$

Dermed har vi vist at k berre er avhengig av X_{+1}, \dots, X_{+t_2} . Av dette følgjer det at (i) og (ii) er oppfylt.

MERK: Om vi slår saman dei $s-t_2$ cellene som ikkje er involvert i problemet (H_1, A_1) og studerer dette problemet under modellen

$$(4.16) \begin{cases} \Pr\left(\bigcap_{i=1}^2 \bigcap_{j=1}^{t_2} (X_{ij} = x_{ij})\right) = \prod_{i=1}^2 \frac{N_i!}{\prod_{j=1}^{t_2} x_{ij}! x_{is_0}!} \prod_{j=1}^{t_2} q_{ij}^{x_{ij}} q_{is_0}^{x_{is_0}}, \\ q_{1j} q_{2s_0} = q_{2j} q_{1s_0} \exp(a_j \gamma), \quad j = 1, 2, \dots, t_2 \end{cases}$$

der a_j -ane er gitt ved (4.13),

$$q_{is_0} = \sum_{j=t_2+1}^s q_{ij} = 1 - \sum_{j=1}^{t_2} q_{ij}, \quad i = 1, 2$$

og

$$x_{is_0} = \sum_{j=t_2+1}^s x_{ij} = N_i - \sum_{j=1}^{t_2} x_{ij}, \quad i = 1, 2,$$

får vi frå kap. 4.1. at testen som forkastar H_1 (dvs. $\gamma=0$) mot A_1'

$$Y = \sum_{j=1}^{t_2} a_j X_{1j} > k(X_{+1}, \dots, X_{+t_2})$$

er O.A.S.S. med nivå ϵ .

A_1' er den delen av A_1 som er avgrensa ved (4.16). Vi har dermed vist at den metoden vi etablerte i den fulle tabellen under modellen gitt ved (4.9) og (4.13), er identisk med den metoden vi har oppnådd i den samanslegne tabellen under modellen (4.16).

5. ASYMPTOTISKE METODAR

I innleiinga peika vi på at dei fleste analysemetodar for kate-
goriske data er baserte på asymptotiske resultat, dvs. metodane er utvikla
under føresetnad av at talet på observasjonar er "stort". Nedafor skal
vi presentere asymptotiske utgåver av metodane gitt i kapitla 4 og 5. Desse
er nyttige når materialet er så stort at dei eksakte metodane blir rekne-
messig for ressurskrevjande.

Dei asymptotiske metodane nedafor er baserte på normaltilnærmingar
av multivariate hypergeometriske eller binomiske fordelingar. Tilnærming-
ane er gyldige når det ikkje er for småe observasjonstal i dei cellene som
inngår i observatoren.

(i) Uavhengigheit (kap. 3.1.)

Vi forkastar hypotesen om uavhengigheit, $H : \gamma = 0$ mot $A : \gamma > 0$
når

$$\frac{Y - \mu_1}{\sigma_1} \geq k_{1-\epsilon}$$

der $k_{1-\epsilon}$ tilfredsstillir $\Phi(k_{1-\epsilon}) = 1-\epsilon$, $\Phi(\cdot)$ er den kumulative fordelings-
funksjon til den standardiserte normalfordelinga,

$$\mu_1 = E_H(Y | X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1})$$

og

$$\sigma_1^2 = \text{var}_H(Y | X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1}).$$

I den betinga fordelinga under H vert X_{ij} hypergeometrisk fordelt. Dette gir

$$\mu_1 = \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} E_H(X_{ij} | X_{i+}, X_{+j}) = \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} \frac{X_{i+} X_{+j}}{N}$$

og

$$\text{var}_H(X_{ij} | X_{i+}, X_{+j}) = \frac{X_{i+} (N - X_{i+}) X_{+j} (N - X_{+j})}{N^2 (N - 1)}.$$

Vi skal no finne den betinga kovariansen mellom X_{ij} og X_{kl} ved å
la

$$X_{ij} = I_{i1} + \dots + I_{ik} X_{i+}$$

og

$$X_{kl} = J_{k1} + \dots + J_{kl} X_{k+}$$

der

$$I_{ig} = \begin{cases} 1 & \text{hvis } g^{\text{te}} \text{ eining trekt frå gruppe } i \text{ er med i kategori} \\ & \text{(kolonne) } j \\ 0 & \text{ellers} \end{cases}$$

og

$$J_{kh} = \begin{cases} 1 & \text{hvis } h^{\text{te}} \text{ eining trekt frå gruppe } k \text{ er med i kategori} \\ & \text{(kolonne) } \ell \\ 0 & \text{ellers} \end{cases}$$

Da blir

$$E(I_{ig} J_{kh}) = \begin{cases} \frac{X_{+j} X_{+\ell}}{N(N-1)} & , (i,g) \neq (k,h) \\ 0 & , (i,g) = (k,h) \end{cases}$$

som vidare gir

$$\text{cov}_H(X_{ij}, X_{kl} | X_{i+}, X_{k+}, X_{+j}, X_{+\ell}) = \begin{cases} \frac{X_{i+} X_{k+} X_{+j} X_{+\ell}}{N^2(N-1)} & , i \neq k \\ & j \neq \ell \\ \frac{(N-X_{i+}) X_{i+} X_{+j} X_{+\ell}}{N^2(N-1)} & , i = k \end{cases}$$

Av symmetrigrunnar får vi

$$\text{cov}_H(X_{ij}, X_{kj} | X_{i+}, X_{k+}, X_{+j}) = - \frac{(N-X_{+j}) X_{+j} X_{i+} X_{k+}}{N^2(N-1)} , i \neq k$$

Dermed får vi at

$$\begin{aligned} \sigma_1^2 &= \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij}^2 \text{var}_H(X_{ij} | X_{i+}, X_{+j}) \\ &+ \sum_{(i,j) \neq (k,\ell)} \sum_{ij, k\ell} a_{ij} a_{k\ell} \text{cov}_H(X_{ij}, X_{k\ell} | X_{i+}, X_{k+}, X_{+j}, X_{+\ell}) \\ &= \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij}^2 \frac{X_{i+} (N-X_{i+}) X_{+j} (N-X_{+j})}{N^2(N-1)} \\ &+ \sum_{i \neq k} \sum_{j \neq \ell} \sum_{ij, k\ell} a_{ij} a_{k\ell} \frac{X_{i+} X_{k+} X_{+j} X_{+\ell}}{N^2(N-1)} + \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} \sum_{\ell=1}^{s-1} a_{ij} a_{i\ell} \frac{(N-X_{i+}) X_{i+} X_{+j} X_{+\ell}}{N^2(N-1)} \end{aligned}$$

$$- \sum_{i=1}^{r-1} \sum_{k=1}^{r-1} \sum_{j=1}^{s-1} \frac{(N-X_{+j})X_{+j}X_{i+}X_{k+}}{N^2(N-1)} a_{ij} a_{kj}.$$

(ii) Symmetri (kap. 3.2)

Vi forkastar hypotesen om symmetri, $H : \gamma = 0$ mot $A : \gamma > 0$ når

$$\frac{Z - \mu_2}{\sigma_2} \geq k_{1-\varepsilon}$$

der $k_{1-\varepsilon}$ er slik at $\Phi(k_{1-\varepsilon}) = 1-\varepsilon$.

Som vi peika på i kap. 3.2., har vi at X_{ij} -ane ($i < j$) er uavhengige og binomisk ($W_{ij}, \frac{1}{2}$) under H når W er gitt. Følgjeleg får vi

$$\mu_2 = E_H(Z|W) = \sum_{i < j} \sum a_{ij} E_H(X_{ij}|W_{ij}) = 1/2 \sum_{i < j} \sum a_{ij} W_{ij}$$

og

$$\sigma_2^2 = \text{var}_H(Z|W) = \sum_{i < j} \sum a_{ij}^2 \text{var}(X_{ij}|W_{ij}) = 1/4 \sum_{i < j} \sum a_{ij}^2 W_{ij}.$$

(iii) Avgrensa symmetri (kap. 3.3.)

Vi forkastar hypotesen om avgrensa symmetri, $H : \gamma = 0$ mot $A : \gamma > 0$ når

$$\frac{Z_1 - \mu_3}{\sigma_3} \geq k_{1-\varepsilon}$$

der $k_{1-\varepsilon}$ er slik at $\Phi(k_{1-\varepsilon}) = 1-\varepsilon$. På tilsvarende måte som for "symmetri" finn vi

$$\mu_3 = E_H(Z_1|W_1) = 1/2 \sum_{\substack{(i,j) \in S \\ i < j}} \sum a_{ij} W_{ij}$$

og

$$\sigma_3^2 = \text{var}_H(Z_1|W_1) = 1/4 \sum_{\substack{(i,j) \in S \\ i < j}} \sum a_{ij}^2 W_{ij}.$$

(iv) Diagonal-parameter symmetri (kap. 3.6.)

Vi forkastar hypotesen om diagonal-parameter symmetri, $H : \gamma = 0$ mot $A : \gamma > 0$ når

$$\frac{Z_4 - \mu_4}{\sigma_4} \geq k_{1-\varepsilon}$$

der $k_{1-\varepsilon}$ er slik at $\Phi(k_{1-\varepsilon}) = 1-\varepsilon$.

I den betingta fordelinga G_4 (kap. 3.6) er X_{2r} , (X_{2r-1}, X_{3r}) ,
 \dots , $(X_{23}, X_{34}, \dots, X_{r-1r})$ uavhengige og (multivariat) hypergeometrisk
fordelte. Dermed blir kvar X_{ii+t} , $i = 2, 3, \dots, r-t$; $t = 1, 2, \dots, r-2$
hypergeometrisk fordelt i den betingta fordelinga under H . Ved å la

$$M_t = \sum_{i=1}^{r-t} W_{ii+t}$$

får vi derfor

$$\begin{aligned} \mu_4 &= E_H(Z_4 | V, W) = \sum_{t=1}^{r-2} \sum_{i=2}^{r-t} a_{it} E_H(X_{ii+t} | V, W) \\ &= \sum_{t=1}^{r-2} (V_t / M_t) \sum_{i=2}^{r-t} a_{it} W_{ii+t} \end{aligned}$$

og

$$\begin{aligned} \sigma_4^2 &= \text{var}_H(Z_4 | V, W) = \sum_{t=1}^{r-2} \text{var} \left(\sum_{i=2}^{r-t} a_{it} X_{ii+t} | V, W \right) \\ &= \sum_{t=1}^{r-2} \left[\sum_{i=2}^{r-t} a_{it}^2 \text{var}_H(X_{ii+t} | V, W) \right. \\ &\quad \left. + \sum_{i \neq k} \sum_{i=2}^{r-t} \sum_{k=2}^{r-t} a_{it} a_{kt} \text{cov}(X_{ii+t}, X_{kk+t} | V, W) \right] \\ &= \sum_{t=1}^{r-2} \left[\sum_{i=2}^{r-t} a_{it}^2 \frac{V_t (M_t - V_t) W_{ii+t} (M_t - W_{ii+t})}{M_t^2 (M_t - 1)} \right. \\ &\quad \left. - \sum_{i \neq k} \sum_{i=2}^{r-t} \sum_{k=2}^{r-t} a_{it} a_{kt} \frac{V_t (M_t - V_t) W_{ii+t} W_{kk+t}}{M_t^2 (M_t - 1)} \right] \end{aligned}$$

(v) Homogenitet (kap. 4.1.)

Dei tilnærma metodane (i), (ii) og (iii) er baserte på at modellen
for tabellen er multinomisk. Prosedyren (i) kan nyttast til homogenitets-
testing når modellen for tabellen er produktmultinomisk. Da blir $X_{i+} = N_i$,
 $i = 1, 2, \dots, s$ ikkje-stokastiske storleiker. Med dei spesielle val av

a_{ij} -ane frå kap. 4.a., blir (i) ein tilnærma metode til å vurdere trend i tabellen.

I kap. 4.3. presenterte vi eksakte metodar for ulike problem i $2 \times s$ -tabellen. Vi får da for den tilnærma varianten til metoden i kap. 4.3. (I); forkast H mot A når

$$\frac{Y - \mu_5}{\sigma_5} \geq k_{1-\epsilon}$$

der $k_{1-\epsilon}$ tilfredsstillir $\Phi(k_{1-\epsilon}) = 1-\epsilon$, $Y = \sum_{j=1}^{s-1} a_j X_{1j}$,

$$\mu_5 = E_H(Y | X_{+1}, \dots, X_{+s-1}) = (N_1/N) \sum_{j=1}^{s-1} a_j X_{+j}$$

og

$$\begin{aligned} \sigma_5^2 &= \text{var}_H(Y | X_{+1}, \dots, X_{+s-1}) \\ &= \sum_{j=1}^{s-1} a_j^2 \text{var}_H(X_{1j} | X_{+j}) + \sum_{j \neq \ell}^{s-1} \sum_{\ell}^{s-1} (\text{cov}_H(X_{1j}, X_{1\ell} | X_{+j}, X_{+\ell})) a_j a_\ell \\ &= \sum_{j=1}^{s-1} a_j^2 \frac{N_1(N-N_1)X_{+j}(N-X_{+j})}{N^2(N-1)} - \sum_{j \neq \ell}^{s-1} \sum_{\ell}^{s-1} \frac{N_1(N-N_1)X_{+j}X_{+\ell}}{N^2(N-1)} a_j a_\ell \end{aligned}$$

Ved å erstatte $s-1$ med t_2 i observatoren

$$\frac{Y - \mu_5}{\sigma_5}$$

blir metoden gitt ved denne observatoren, den tilnærma varianten til den eksakte metoden i kap. 4.3. (II).

Fordelinga til dei ulike metodeobservatorane i kapitla 3 og 4 er diskrete. Ovafor har vi presentert kontinuerlege tilnærmingar til dei diskrete fordelingane. Bruk av kontinuerlege tilnærmingar vil gi opphav til feil; spesielt når utvalet er av moderat eller liten storleik. Ved å nytte kontinuitetskorreksjonar kan vi til ein viss grad redusere den feilen som har sin årsak i at den diskrete fordelinga er blitt erstatta av ei kontinuerleg fordeling.

La T representere den typen observatorer som blir omtala i kapitla 3 og 4. La vidare μ vere den aktuelle betinga forventninga og σ^2 den aktuelle betinga variansen. Vi føreset at dei moglege verdiane til T er ekvidistante, dvs. to nabovertiar har same avstand λ .

Ved å innføre kontinuitetskorreksjonar, dvs. vi forkastar når

$$\frac{T - \mu - \lambda/2}{(\sigma^2 - \lambda^2/12)^{1/2}} \geq k_{1-\varepsilon},$$

oppnår vi betring av presisjonen til dei tilnærma metodane. Hvis a_{ij} -ane er valt på ein slik måte at dei moglege verdiane til observatorane (i), (ii), (iii) eller (iv) er ekvidistante, bør vi altså innføre kontinuitetskorreksjonar. I mange situasjonar vil det vere naturleg å velje a_{ij} -ane lik 1 og -1. Da vert dei moglege verdiane til observatorane ekvidistante med avstand $\lambda = 1$. Hajek & Sidak (1967) gir ein omtale av kontinuitetskorreksjonar.

6. SENTRALE OMGREP I TO-VEGSTABELLEN UTTRYKT VED PARAMETRANE I DEN LOG-LINEÆRE MODELLEN

Log-lineær analyse av tabellar er handsama i Haldorsen, T (1977b). Når det gjeld inferensproblem, tolkning av parametrane, osv. i log-lineære modellar, så viser vi til denne publikasjonen. I dette kapitlet skal vi nøye oss med å omtale korleis dei ulike omgrepa (problema) frå kap. 3 og kap. 4 kan uttrykkast ved parametrane i den log-lineære modellen.

Den log-lineære modellen er definert ved

$$\log p_{ij} = u + u_1(i) + u_2(j) + u_{12}(i,j), \quad i = 1, 2, \dots, r; \\ j = 1, 2, \dots, s$$

der

$$\sum_i u_1(i) = \sum_j u_2(j) = \sum_i u_{12}(i,j) = \sum_j u_{12}(i,j) = 0.$$

u gir uttrykk for gjennomsnittseffekt. u_1 og u_2 kan vi tolke som hovud-effektar for respektive variabel A og variabel B, medan u_{12} uttrykker samspeleffekten mellom A og B.

Som i kap. 3 ser vi på observasjonane i rxs-tabellen som resultat av ei multinomisk forsøksrekke. Vi får da for parametrane i den log-lineære modellen

$$u = \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s \log p_{ij} = \log \bar{p}_{..},$$

$$u_1(i) = \frac{1}{s} \sum_{j=1}^s \log p_{ij} - u = \log \frac{\bar{p}_{i.}}{\bar{p}_{..}},$$

$$u_2(j) = \frac{1}{r} \sum_{i=1}^r \log p_{ij} - u = \log \frac{\bar{p}_{.j}}{\bar{p}_{..}},$$

$$u_{12}(i,j) = \log p_{ij} - u_1(i) - u_2(j) - u = \log \frac{\bar{p}_{ij} \bar{p}_{..}}{\bar{p}_{i.} \bar{p}_{.j}}$$

der

$$\bar{p}_{..} = \left(\sum_{i=1}^r \sum_{j=1}^s p_{ij} \right)^{1/rs}, \quad \bar{p}_{i.} = \left(\sum_{j=1}^s p_{ij} \right)^{1/s}, \quad \bar{p}_{.j} = \left(\sum_{i=1}^r p_{ij} \right)^{1/r}$$

og

$$p_{ij} = \Pr(A_i \cap B_j).$$

Ved å nytte definisjonane i kap. 3 får vi

(i) uavhengigheit viss og berre viss $u_{12}(i,j) = 0$ for alle i og j

(ii) symmetri viss og berre viss $u_1(i) + u_2(j) + u_{12}(i,j) = u_1(j) + u_2(i) + u_{12}(j,i)$, $i \neq j$.

I ein modell med symmetri har vi derfor

$$u_1(i) = u_2(i) \quad \text{og} \quad u_{12}(i,j) = u_{21}(j,i), \quad i \neq j.$$

(iii) avgrensa symmetri viss og berre viss

$$u_1(i) + u_2(j) + u_{12}(i,j) = u_1(j) + u_2(i) + u_{12}(j,i),$$

$$(i,j) \in S \text{ der } S \subseteq \{(i,j) : i \neq j\}.$$

(iv) samspelsymmetri viss og berre vis

$$u_{12}(i,j) = u_{12}(j,i), \quad i \neq j.$$

(v) marginal homogenitet viss og berre viss

$$\begin{aligned} & (\exp u_1(i)) \sum_{k=1}^r \exp(u_2(k) + u_{12}(i,k)) \\ &= (\exp u_2(i)) \sum_{k=1}^r \exp(u_1(k) + u_{12}(k,i)), \quad i = 1, 2, \dots, r. \end{aligned}$$

(vi) diagonal-parametersymmetri viss og berre viss

$$u_1(i) - u_2(i) + u_2(j) - u_1(j) + u_{12}(i,j) - u_{12}(j,i) = \xi_t,$$

$$t = j-i, \quad i \neq j.$$

Når modellen for tabellen er produktmultinomisk (kap. 4.), får vi at den log-lineære formuleringa av denne modellen svarar til den ovafor med endra rammekrav. Vi kan derfor lett uttrykke dei ulike problema i kap. 4. ved parametrane u' , $u'_1(i)$, $u'_2(j)$ og $u'_{12}(i,j)$ i den log-lineære modellen. Homogenitet i tabellen vil seie at samspelleddet $u'_{12}(i,j)$ er lik 0. Når det gjeld hypotesen (H_1) i kap. 4.3.(II), får vi følgjande samanheng mellom parametrane i den log-lineære modellen;

$$u'_{12}(1,j) = u'_1(2), \quad j = 1, 2, \dots, t_2.$$

Dei mest nytta metodane som er utvikla for analyse av parametrane i log-lineære modellar er asymptotiske og dessutan avgrensa til å gjelde for hierarkiske modellar (sjå Haldorsen (1977b)). Vi kan derfor ikkje nytte ECTA (program som er i bruk i Byrået) til f.eks. å studere samspelet u_{12} under ein modell der ein eller begge hovudeffektane er lik 0, medan dette let seg gjere ved hjelp av resultatata frå kap. 2.

I kap. 6, Aaberge (1979) peika vi på at

$$p_{11} = p_{22} \quad \text{og} \quad p_{12} = p_{21}$$

er ekvivalent med at begge hovudeffektane i den log-lineære formuleringa for 2x2-tabellen er lik 0; dvs. vi har ein modell med berre samspel. I kap. 5, Aaberge (1979) nemnde vi eit eksempel på ein tabell, der vi kunne nytta ein slik modell. Sidan føresetnaden om $p_{11} = p_{22}$ i dette eksemplet var heller tvilsam, skal vi nedafor gi eit meir realistisk eksempel på ein slik situasjon.

Vi vil undersøke om det er samanheng mellom kjønn og første-fødd av ein-egga tvillingar. Vi kan basere undersøkinga på eit tilfeldig utval av ein-egga tvillingar. Modellen for tabellen blir multinomisk med parametrar

		Fødd sist		
		Jente	Gut	Sum
Fødd først	Jente	p_{11}	p_{12}	p_{1+}
	Gut	p_{21}	p_{22}	p_{2+}
	Sum	p_{+1}	p_{+2}	1

I denne situasjonen er både

$$p_{12} = p_{21} \quad \text{og} \quad p_{11} = p_{22}$$

rimelege føresetnader, dvs. begge hovudeffektane i den log-lineære modell-formuleringa er lik 0.

KONSTRUKSJON AV OVERALT STERKASTE STYRKERETTE TESTAR UNDER EIN POISSON-MODELL

I kapitla 2 og 3 såg vi på to-vegstabellen som resultat av ei multinomisk forsøksrekke. Utvalstorleiken N var fast. I Statistisk Sentralbyrå blir det nytta register for trekking av utval. Utvalstorleiken er derfor fastsett på førehand. Innafor mange forskningsfelt har ein ikkje denne føremonen. Observasjonane blir derfor ofte henta inn på ein slik måte at også den totale utvalstorleiken er ein stokastisk variabel. Hvis vi plukkar ut ein tabell frå eit større materiale, vil det ofte gi opphav til ein Poisson-modell for tabellen. Slike situasjonar kan også oppstå i Statistisk Sentralbyrå. Vi ser da på observasjonane frå kvar av cellene i $r \times s$ -tabellen som uavhengige Poisson-fordelte variable. Vi får altså tabellen

	B_1	B_2	...	B_j	...	B_s	Sum
A_1	X_{11}	X_{12}		X_{1j}		X_{1s}	X_{1+}
A_2	X_{21}	X_{22}		X_{2j}		X_{2s}	X_{2+}
...							
A_i	X_{i1}	X_{i2}		X_{ij}		X_{is}	X_{i+}
...							
A_r	X_{r1}	X_{r2}		X_{rj}		X_{rs}	X_{r+}
Sum	X_{+1}	X_{+2}		X_{+j}		X_{+s}	X_{++}

der $X_{ij} \sim \text{Poisson}(\lambda_{ij})$, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$.

λ_{ij} blir sett på som eit mål for den intensitet $A_i \cap B_j$ opptre med. Vi får da for simultanfordelinga (punktsannsynet) til X_{ij} -ane

$$(1) \quad \Pr\left(\bigcap_{i,j} (X_{ij} = x_{ij})\right) = \prod_{i=1}^r \prod_{j=1}^s \left(\frac{\lambda_{ij}^{x_{ij}}}{x_{ij}!} e^{-\lambda_{ij}} \right).$$

Modell (2.1) i kap. 2 kan vi sjå på som ein betinga Poisson-modell gitt X_{++} . Vi finn da for

$$p_{ij} = \Pr(A_i \cap B_j)$$

uttrykt ved parametrane i Poisson-modellen (1)

$$p_{ij} = \frac{\lambda_{ij}}{\lambda_{++}}, \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, s$$

der

$$\lambda_{++} = \sum_{i=1}^r \sum_{j=1}^s \lambda_{ij}$$

La \mathcal{P} vere fordelingsklassen definert ved (1). \mathcal{P} er eksponensiell sidan vi for einkvar $P \in \mathcal{P}$ har at

$$(2) \quad dP = (\exp(x_{++} - \lambda_{++})) (\exp(\sum_{i=1}^r \sum_{j=1}^s x_{ij} \log \lambda_{ij})) dP_0$$

der P_0 er sannsynfordelinga som svarar til $\lambda_{ij} = 1$, $i = 1, 2, \dots, r$; $j = 1, 2, \dots, s$.

La

$$(3) \quad \theta_{ij} = \log \lambda_{ij}, \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, s$$

dvs.

$$\lambda_{ij} = \exp(\theta_{ij}), \quad i = 1, 2, \dots, r; \quad j = 1, 2, \dots, s.$$

La $\Theta = \theta(\mathcal{P})$, $\theta = (\theta_{11}, \dots, \theta_{1s}, \theta_{21}, \dots, \theta_{2s}, \dots, \theta_{r1}, \dots, \theta_{rs})'$.

Transformasjonen

$$P \rightarrow \theta(P), \quad P \in \mathcal{P}$$

er ein-eintydig og $\dim \Theta = \dim \mathcal{P} = rs$. Dette gir at parametriseringa

$$\{P_\theta : \theta \in \Theta\}$$

gitt ved

$$(4) \quad dP_\theta = [a(\theta) \exp(\sum_{i=1}^r \sum_{j=1}^s \theta_{ij} x_{ij})] dP_0$$

der $a(\theta) = \exp(x_{++} - \sum_{i=1}^r \sum_{j=1}^s \exp(\theta_{ij}))$,

er minimalkanonisk representasjon av \mathcal{P} .

Sidan $\mathcal{M} = \mathbb{R}^N$ er \mathcal{M} ei open mengd. Vi får derfor at den eksponensielle fordelingsklassen \mathcal{P} er regulær. Som i kap. 2 skal vi avgrense oss til å undersøke regulære hypotesar \mathcal{P}_0 . Vidare ønskjer vi å studere problem som er definert ved ein parameter, γ . Dette kan vi gjere ved å studere γ under følgjande restriksjonar av parametrane i Poisson-modellen for tabellen

$$(5) \quad U\left(\frac{Y}{\xi}\right) + v = B\theta$$

der $\gamma \in \mathbb{R}$, $\theta \in \mathcal{M}$, ξ er ein vektor med dimensjon g_0 , B er ei $d_1 \times r$ -matrise med rang d_1 , $d_1 = 1, 2, \dots, r$; U er ei $d_1 \times (d_0 + 1)$ -matrise med rang $d_0 + 1$, $d_0 = 0, 1, \dots, d_1 - 1$, $U \neq 0$, v er ein vektor med dimensjon d_1 . Elementa til ξ er ukjend parametrar, medan elementa til B , U og v er faste reelle tal.

Dette vil seie at

$$\mathcal{P}_0 = \mathcal{P}_{\gamma, B}; \gamma \in \mathbb{H}$$

der $\mathcal{P}_{\gamma, B}$ er definert ved

$$\mathcal{P}_{\gamma, B} = \{P_\theta : B\theta = v + U\left(\frac{Y}{\xi}\right)\}$$

er affin og dermed regulær. Teorem 3 (kap. 4), Lehmann (1959) gir derfor at det eksisterer O.A.S.S. testar med nivå ε for å teste H_1, H_2, H_3 og H_4 mot respektive A_1, A_2, A_3 og A_4 . H_j og A_j ; $j = 1, 2, 3, 4$ er definert i kap. 2.

Vi skal no etablere testar for hypotesen $\gamma = 0$ mot alternativa $\gamma > 0$ dvs. problem av typen H_1, A_1 . Ved å setje inn for $x_{rs} = x_{++} - \sum_{(i,j) \neq (r,s)} x_{ij}$ og (5) i (4) får vi

$$(6) \quad dP_\theta = \left[a(\theta) \exp(\gamma(\theta)T(x) + \sum_{i=1}^{rs-d} \tau_i(\theta)Y_i(x)) \right] dP_0$$

$$\text{der } d = d_1 - d_0.$$

Funksjonsformene til $\gamma, \tau_1, \tau_2, \dots, \tau_{rs-d}$ og dermed til $T, Y_1, Y_2, \dots, Y_{rs-d}$ kjem fram ved fastsetjing av B, U og v . Analogt med Setning 2, gir Teorem 3 (kap. 4) Lehmann (1959),

Setning 4. La X ha fordeling gitt ved (4) og (5). For å teste $H : \gamma = 0$ mot $A : \gamma > 0$ så eksisterer det ein overalt sterkast styrkerett (O.A.S.S.)-test δ med nivå ε . Testen er definert ved

$$\delta(x) = \begin{cases} 1 & T(x) > k(y_1, y_2, \dots, y_{rs-d}) \\ \mu & T(x) = k(y_1, y_2, \dots, y_{rs-d}) \\ 0 & T(x) < k(y_1, y_2, \dots, y_{rs-d}) \end{cases}$$

der k og μ er gitt ved

$$E_H(\delta(X) | y_1, y_2, \dots, y_{rs-d}) = \epsilon.$$

Om vi ønskjer å studere uavhengigheit i tabellen under modellen gitt ved (1) og (5) (dvs. modellen gitt ved (6)), kan vi gjere dette ved å teste hypotesen $\gamma = 0$ mot alternativa $\gamma > 0$, der γ er gitt ved

$$\log \frac{\lambda_{ij}^{\lambda_{rs}}}{\lambda_{is}^{\lambda_{rj}}} = a_{ij}\gamma, \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1$$

dvs.

$$(7) \quad \theta_{ij} - \theta_{is} - \theta_{rj} + \theta_{rs} = a_{ij}\gamma, \quad i = 1, 2, \dots, r-1; \quad j = 1, 2, \dots, s-1.$$

a_{ij} -ane er fastlagde reelle tal. Vi har derfor uavhengigheit i rxs-tabellen hvis og berre hvis $\gamma = 0$. Dessutan er (7) opplagt med i (5).

Ved å setje inn for (7) i (4) får vi for (6)

$$(8) \quad dP_{\theta} = \left[a(\theta) \exp\left(\gamma \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} + \sum_{i=1}^r \theta_{is} x_{i+} + \sum_{j=1}^s \theta_{rj} x_{+j} - \theta_{rs} x_{++}\right) \right] dP_0$$

der

$$x_{r+} = x_{++} - \sum_{i=1}^{r-1} x_{i+}, \quad x_{+s} = x_{++} - \sum_{j=1}^{s-1} x_{+j}.$$

Da gir Setning 4 at testen som forkastar

$$H : \gamma = 0 \text{ mot } A : \gamma > 0$$

når

$$Y = \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} X_{ij} > k(X_{1+}, \dots, X_{r-1+}, X_{+1}, \dots, X_{+s-1}, X_{++})$$

er O.A.S.S. test med nivå ϵ .

På tilsvarende måte som i kap. 3.1. finn vi at k vert fastsett slik at

$$\sum_{\ell > k} g(\ell | x_{1+}, \dots, x_{r-1+}, x_{+1}, \dots, x_{+s-1}, x_{++}) \\ + \mu g(k | x_{1+}, \dots, x_{r-1+}, x_{+1}, \dots, x_{+s-1}, x_{++}) = \varepsilon$$

der $0 \leq \mu \leq 1$,

$$g(y | x_{1+}, \dots, x_{r-1+}, x_{+1}, \dots, x_{+s-1}, x_{++}) \\ = \sum_{x \in S_1} \frac{\binom{x_{1+}}{x_{11}, \dots, x_{1s}} \binom{x_{2+}}{x_{21}, \dots, x_{2s}} \cdots \binom{x_{r+}}{x_{r1}, \dots, x_{rs}}}{\binom{x_{++}}{x_{+1}, \dots, x_{+s}}}$$

$$S_1 = \{x : \sum_{i=1}^{r-1} \sum_{j=1}^{s-1} a_{ij} x_{ij} = y\}, \quad x = (x_{11}, \dots, x_{1s}, \dots, x_{r1}, \dots, x_{rs})'$$

$$x_{rj} = x_{+j} - \sum_{i=1}^{r-1} x_{ij}, \quad j = 1, 2, \dots, s \quad \text{og}$$

$$x_{is} = x_{i+} - \sum_{j=1}^{s-1} x_{ij}, \quad i = 1, 2, \dots, r.$$

Prosedyren blir altså den same som den vi utvikla for uavhengigheit i kap. 3.1. For å studere dei andre problema frå kap. 3 under modellar gitt ved (1) og (5), får vi tilsvarande at dei respektive prosedyrane frå kap. 3 kan nyttast.

NULLFORDELINGA TIL GENERALISERTE FISHER-IRWINS TEST

Vi vil nedafor bruke same notasjon som i kap. 3.1.

Det følgjer opplagt at den betingta fordelinga for Y gitt (X_{1+}, \dots) er gitt ved

$$g_1(y | x_{1+}, \dots, x_{r-1+}, x_{+1}, \dots, x_{+s-1}) \\ = \sum_{x \in S_1} \left(\prod_{i=1}^r \prod_{j=1}^s x_{ij}! \right)^{-1} \left[\sum_{x \in S_2} \left(\prod_{i=1}^r \prod_{j=1}^s x_{ij}! \right)^{-1} \right]^{-1}$$

der $S_2 = \{x : \sum_{j=1}^s x_{ij} = x_{i+}, i = 1, 2, \dots, r; \sum_{i=1}^r x_{ij} = x_{+j}, j = 1, 2, \dots, s\}$

Ved å nytte Lemma 2 med $t = r-1$ og $u = s-1$ får vi

$$g_1(y | x_{1+}, \dots, x_{r-1+}, x_{+1}, \dots; x_{+s-1}) \\ = \sum_{x \in S_1} \left(\prod_{i=1}^r \frac{x_{i+}!}{x_{i1}! \dots x_{is}!} \right) \left(\frac{N!}{x_{+1}! \dots x_{+s}!} \right)^{-1} \\ = \sum_{x \in S_1} \frac{\binom{x_{1+}}{x_{11}, \dots, x_{1s}} \binom{x_{2+}}{x_{21}, \dots, x_{2s}} \dots \binom{x_{r+}}{x_{r1}, \dots, x_{rs}}}{\binom{N}{x_{+1}, \dots, x_{+s}}}$$

Lemma 2. La $x = (x_{11}, \dots, x_{1u+1}, x_{21}, \dots, x_{2u+1}, \dots, x_{t+1}, \dots, x_{t+u+1})'$,
 $\sum_{i=1}^{t+1} \sum_{j=1}^{u+1} x_{ij} = N$ og $S = \{x : \sum_{j=1}^{u+1} x_{ij} = x_{i+}, i=1, 2, \dots, t+1; \sum_{i=1}^{t+1} x_{ij} = x_{+j}, j=1, 2, \dots, u+1\}$

Da har vi

$$\sum_{x \in S} \left(\prod_{i=1}^{t+1} \prod_{j=1}^{u+1} x_{ij}! \right)^{-1} = N! \left[\left(\prod_{i=1}^{t+1} x_{i+}! \right) \left(\prod_{j=1}^{u+1} x_{+j}! \right) \right]^{-1}$$

Prov. Ved å omforme uttrykket innafor summeteiknet får vi

$$\begin{aligned}
 \sum_{x \in S} \left(\prod_{i=1}^{t+1} \prod_{j=1}^{u+1} x_{ij} \right)^{-1} &= \sum_{x \in S} \frac{1}{N!} \binom{N}{x_{1+}} \left[\binom{x_{1+}}{x_{11}} \binom{N-x_{1+}}{x_{+1}-x_{11}} \right. \\
 &\cdot \left. \binom{x_{1+}-x_{11}}{x_{12}} \binom{N-x_{1+}-x_{+1}+x_{11}}{x_{+2}-x_{12}} \dots \binom{x_{1+}-\sum_{j=1}^{u-1} x_{1j}}{x_{1u}} \binom{N-x_{1+}-\sum_{j=1}^{u-1} (x_{+j}-x_{1j})}{x_{+u}-x_{1u}} \right] \\
 &\cdot \left[\binom{x_{+1}-x_{11}}{x_{21}} \binom{x_{+2}-x_{12}}{x_{22}} \dots \binom{x_{+u}-x_{1u}}{x_{2u}} \binom{N-x_{1+}-\sum_{j=1}^u (x_{+j}-x_{1j})}{x_{2+}-\sum_{j=1}^u x_{2j}} \right] \\
 &\cdot \left[\binom{x_{+1}-x_{11}-x_{21}}{x_{31}} \binom{x_{+2}-x_{12}-x_{22}}{x_{32}} \dots \binom{x_{+u}-x_{1u}-x_{2u}}{x_{3u}} \binom{N-x_{1+}-x_{2+}-\sum_{j=1}^u (x_{+j}-x_{1j}-x_{2j})}{x_{3+}-\sum_{j=1}^u x_{3j}} \right] \\
 &\dots \dots \dots \left[\binom{x_{+1}-\sum_{i=1}^{t-1} x_{i1}}{x_{t1}} \binom{x_{+2}-\sum_{i=1}^{t-1} x_{i2}}{x_{t2}} \dots \dots \dots \right. \\
 &\dots \dots \dots \left. \binom{x_{+u}-\sum_{i=1}^{t-1} x_{iu}}{x_{tu}} \binom{N-\sum_{i=1}^{t-1} x_{i+}-\sum_{j=1}^u (x_{+j}-\sum_{i=1}^{t-1} x_{ij})}{x_{t+}-\sum_{j=1}^u x_{tj}} \right] \\
 &= \frac{1}{N!} \binom{N}{x_{1+}} \left[\sum \binom{x_{+1}}{x_{11}} \binom{N-x_{1+}}{x_{+1}-x_{11}} \sum \binom{x_{1+}-x_{11}}{x_{12}} \binom{N-x_{1+}-x_{+1}+x_{11}}{x_{+2}-x_{12}} \dots \dots \dots \right. \\
 &\dots \dots \dots \sum \binom{x_{1+}-\sum_{j=1}^{u-1} x_{1j}}{x_{1u}} \binom{N-x_{1+}-\sum_{j=1}^{u-1} (x_{+j}-x_{1j})}{x_{+u}-x_{1u}} \sum \binom{x_{+1}-x_{11}}{x_{21}} \sum \binom{x_{+2}-x_{12}}{x_{22}} \\
 &\dots \dots \dots \sum \binom{x_{+u}-x_{1u}}{x_{2u}} \binom{N-x_{1+}-\sum_{j=1}^u (x_{+j}-x_{1j})}{x_{2+}-\sum_{j=1}^u x_{2j}} \sum \binom{x_{1+}-x_{11}-x_{21}}{x_{31}} \sum \binom{x_{1+}-x_{11}-x_{21}}{x_{31}}
 \end{aligned}$$

$$\cdot \sum_{x_{32}}^{x_{+2}-x_{12}-x_{22}} \dots \sum_{x_{3u}}^{x_{+u}-x_{1u}-x_{2u}} \binom{N-x_{1+}-x_{2+}-\sum_{j=1}^u (x_{+j}-x_{1j}-x_{2j})}{x_{3+}-\sum_{j=1}^u x_{3j}}$$

$$\dots \sum_{x_{t1}}^{x_{+1}-\sum_{i=1}^{t-1} x_{i1}} \sum_{x_{t2}}^{x_{+2}-\sum_{i=1}^{t-1} x_{i2}} \dots$$

$$\cdot \sum_{x_{tu}}^{x_{+u}-\sum_{i=1}^{t-1} x_{iu}} \binom{N-\sum_{i=1}^{t-1} x_{i+}-\sum_{j=1}^u (x_{+j}-\sum_{i=1}^{t-1} x_{ij})}{x_{t+}-\sum_{j=1}^u x_{tj}}$$

$$= \frac{1}{N!} \binom{N}{x_{+1}} \binom{N-x_{+1}}{x_{+2}} \dots \binom{N-\sum_{j=1}^{u-1} x_{+j}}{x_{+u}}$$

$$\cdot \binom{N}{x_{1+}} \binom{N-x_{1+}}{x_{2+}} \binom{N-x_{1+}-x_{2+}}{x_{3+}} \dots \binom{N-\sum_{i=1}^{t-1} x_{i+}}{x_{t+}}$$

$$= N! \left[\binom{t+1}{i=1}^{x_{i+}} \binom{u+1}{j=1}^{x_{+j}} \right]^{-1}$$

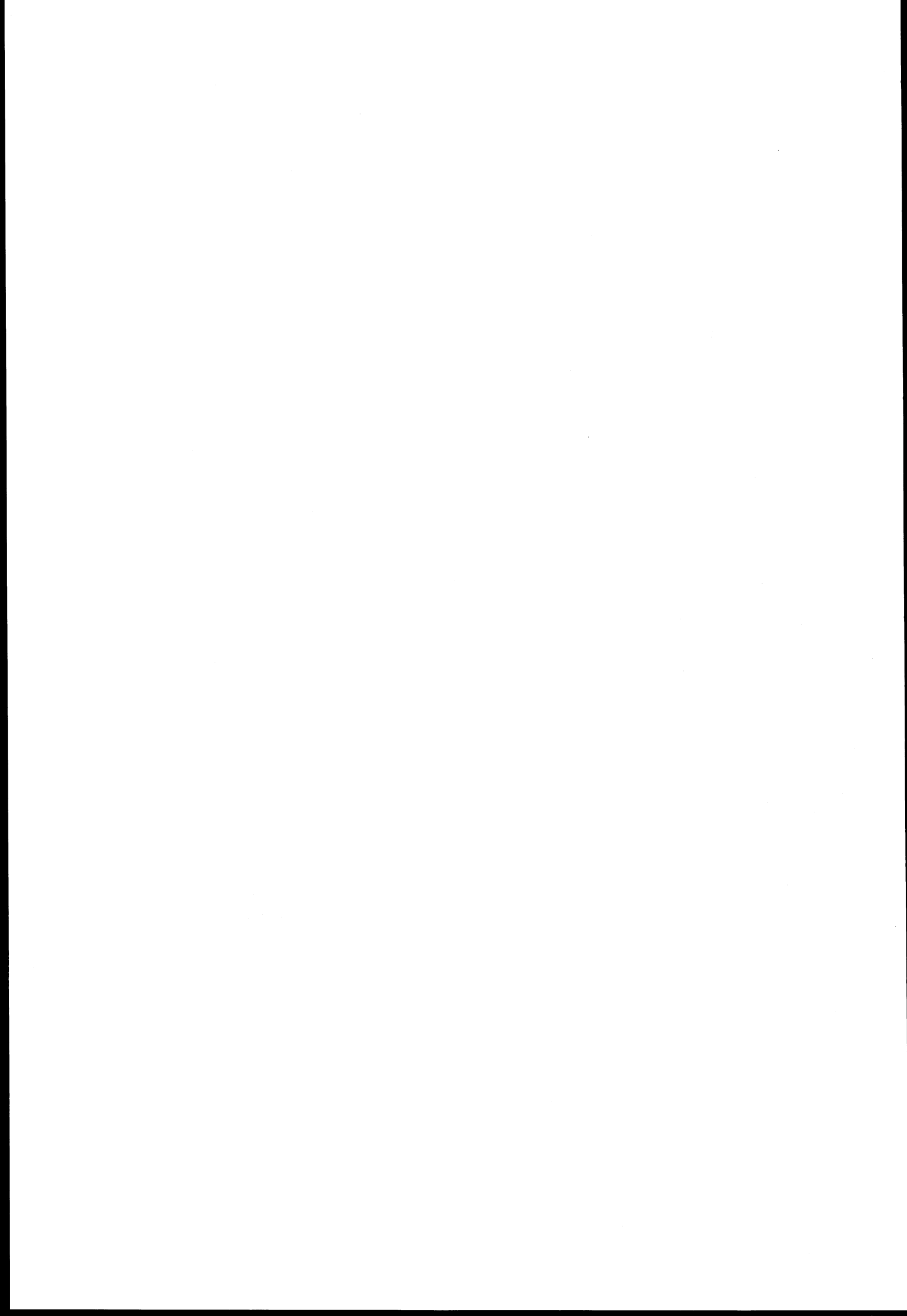
Vi har gjort bruk av følgjande kjende resultat,

$$\sum_{x_{1\ell}}^{x_{1+}-\sum_{j=1}^{\ell-1} x_{1j}} \binom{N-x_{1+}-\sum_{j=1}^{\ell-1} (x_{+j}-x_{1j})}{x_{+\ell}-x_{1\ell}} = \binom{N-\sum_{j=1}^{\ell-1} x_{+j}}{x_{+\ell}}$$

og

$$\Sigma \begin{pmatrix} x_{+1} - \sum_{i=1}^{k-1} x_{i1} \\ x_{k1} \end{pmatrix} \Sigma \begin{pmatrix} x_{+2} - \sum_{i=1}^{k-1} x_{i2} \\ x_{k2} \end{pmatrix} \dots$$

$$\dots \Sigma \begin{pmatrix} x_{+u} - \sum_{i=1}^{k-1} x_{iu} \\ x_{ku} \end{pmatrix} \begin{pmatrix} N - \sum_{i=1}^{k-1} x_{i+} - \sum_{j=1}^u (x_{+j} - \sum_{i=1}^{k-1} x_{ij}) \\ x_{k+} - \sum_{j=1}^u x_{kj} \end{pmatrix} = \begin{pmatrix} N - \sum_{i=1}^{k-1} x_{i+} \\ x_{k+} \end{pmatrix}.$$



Utkome i serien Rapportar fra Statistisk Sentralbyrå (RAPP) - ISSN 0332-8422

Trykt 1980

- " 80/1 Svein Longva, Lorents Lorentsen and Øystein Olsen: Energy in a Multi-Sectoral Growth Model Energi i en flersektors vekstmodell ISBN 82-537-1082-8
- " 80/2 Viggo Jean-Hansen: Totalregnskap for fiske- og fangstnæringen 1975 - 1978 ISBN 82-537-1080-1
- " 80/3 Erik Biørn og Hans Erik Fosby: Kvartalsserier for brukerpriser på realkapital i norske produksjonssektorer ISBN 82-537-1087-9
- " 80/4 Erik Biørn and Eilev S. Jansen: Consumer Demand in Norwegian Households 1973 - 1977 A Data Base for Micro-Econometrics ISBN 82-537-1086-0
- " 80/5 Ole K. Hovland: Skattemodellen LOTTE Testing av framskrivingsmetoder ISBN 82-537-1088-7
- " 80/6 Fylkesvise elektrisitetsprognoser for 1985 og 1990 En metodestudie ISBN 82-537-1091-7
- " 80/7 Analyse av utviklingen i elektrisitetsforbruket 1978 og første halvår 1979 ISBN 82-537-1129-8
- " 80/8 Øyvind Lone: Hovedklassifiseringa i arealregnskapet ISBN 82-537-1104-2
- " 80/9 Tor Bjerkedal: Yrke og fødsel En undersøkelse over betydningen av kvinners yrkesaktivitet for opp treden av fosterskader Occupation and Outcome of Pregnancy ISBN 82-537-1111-5
- " 80/10 Statistikk fra det økonomiske og medisinske informasjonssystem Alminnelige somatiske sykehus 1978 ISBN 82-537-1119-0
- 80/12 Torgeir Melien: Ressursregnskap for jern ISBN 82-537-1138-7
- 80/14 Petter Frenger: Import Share Functions in Input - Output Analysis Importandelsfunksjoner i kryssløpsmodeller ISBN 82-537-1143-3
- 80/15 Den statistiske behandlingen av oljevirkksomheten ISBN 82-537-1150-6
- 80/16 Ådne Cappelen, Eva Ivås og Paal Sand: MODIS IV Detaljerte virkningstabeller for 1978 ISBN 82-537-1142-5
- 80/18 Susan Lingsom: Dagbøker med og uten faste tidsintervaller: En sammenlikning basert på prøveundersøkelse om tidsnytting 1979 Open and Fixed Interval Time Diaries: A Comparison Based on a Pilot Study on Time Use 1979 ISBN 82-537-1158-1
- 80/19 Sigurd Høst og Trygve Solheim: Radio- og fjernsynsundersøkelsen januar - februar 1980 ISBN 82-537-1155-7
- 80/21 Olav Bjerkholt og Øystein Olsen: Optimal kapasitet og fastkraftpotensial i et vannkraftsystem ISBN 82-537-1154-9
- 80/22 Eksakte metodar for analyse av to-vegstabellar ISBN 82-537-1161-1

Pris kr 11,00

Publikasjonen utgis i kommisjon hos H. Aschehoug & Co. og
Universitetsforlaget, Oslo, og er til salgs hos alle bokhandlere.

ISBN 82-537-1161-1
ISSN 0332-8422