Arbeidsnotater

STATISTISK SENTRALBYRÅ

OSLO: Postboks 8131 Dep, Oslo 1
Tlf. (02) *41 38 20

KONGSVINGER: Postboks 510, Stasjonssida, 2201 Kongsvinger
Tlf. (066) *14 988

IO 78/21

6. september 1978

ANALYSE AV TIDSREKKEDATA FOR FØRSTEGANGSGIFTEMÅL PÅ GRUNNLAG AV EN TOKJØNNSMODELL

av

John Dagsvik*)

Innhold

*) Jeg takker Jan Mønnesland og Per Sevaldson for nyttige kommentarer.

1. Innledning

En vesentlig svakhet ved formell demografisk teori er mangelen på operasjonaliserbare teknikker som tar hensyn til avhengigheten mellom kjønnene. De klassiske modellene behandler hvert kjønn separat og en kan dermed risikere å få inkonsistente resultater når modellene skal benyttes i prediksjonsøyemed. Det sentrale problemet synes å være spesifikasjonen av giftemålsfunksjonen - dvs. antallet som gifter seg i hver alder som funksjon av bestanden av ugifte i de ulike aldre. I litteraturen er det foreslått flere matematiske modeller men ingen av dem synes å være helt tilfredsstillende. Det arbeides også med å utvikle numeriske prosedyrer men resultatene er foreløpig sparsomme.

I Byrået er det nylig etablert en historisk datafil med tidsrekkedata over alderspesifikke giftemålsrater (førstegangsgiftemål, kohortfordelinger og periodefordelinger). Dette datamaterialet gjør det ønskelig å ta i bruk analyseteknikker som kan ta hensyn til kompliserte avhengighetstrukturer i materialet.

Det er ønskelig å studere betydningen av endringer i "bestandsvektorene" av ugifte for giftemålsstrukturen, samtidig som en realistisk modell i tillegg bør kunne beskrive eventuelle ikke-forklarte stokastiske variasjoner.

Vi skal her beskrive en modell som har egenskapene ovenfor. Vi tar utgangspunkt i at begivenheten "førstegangsgiftemål" bare inntreffer en gang for hver person dersom den inntreffer. Vi betrakter settet av alderspesifikke giftemålssannsynligheter for hver kohort som realisasjoner av multinomiske sannsynligheter. Disse sannsynlighetene tenkes generert av en fordelingsklasse kalt den multivariate beta-fordeling.

Ved hvert tidspunkt knytter vi kohortene av menn og kvinner sammen ved å la parametrene i fordelingsklassen avhenge av bestandene av ugifte i de ulike kohortene ved det aktuelle tidspunkt. Denne angrepsmåten leder til en simultan modell for alle kohortene hvor bestandene av ugifte og antallet som gifter seg ved hvert tidspunkt inngår. Videre antas det at den ikke-forklarte variasjonen i giftemålsratene utvikler seg i følge en autoregressiv prosess slik at modellen tar hensyn til eventuell uforklart (autokorrelert) variasjon.

I denne behandlingen vil vi se bort fra avgang som skjer før første giftemål (dødsfall, utvandring, m.v.)

2. En dobbeltstokastisk modell

Vi forutsetter at vi har et datasett som består av alderspesifikke giftemålsrater for et visst antall kohorter. La $p_t(x)$ være sannsynligheten for at en person som tilhører kohort t skal gifte seg første gang i alder x. $p_t(x)$, x = 0,16,... er altså sannsynligheten i en multinomisk modell hvor begivenheten "å gifte seg første gang" skjer i alder x eller ikke skjer i det hele tatt. La $p_t(0)$ være sannsynlighetene for at personen ikke gifter seg. La $Y_{tj}(x)$ være lik 1 dersom j-te person i t-te kohort gifter seg i alder x og null ellers. $Y_t(x) = \sum_j Y_{tj}(x)$ er da antallet i kohorten som gifter seg i alder x. Bestanden i kohort t betegnes med N_t og $N_t - \sum_{x \geq 16} Y_t(x)$ betegnes med $Y_t(0)$. Under forutsetning om stokastisk uavhengighet mellom individene blir $\underset{\sim}{Y}_t = (Y_t(x), x=0,\ldots)$ multinomisk fordelt. Dette gjelder innen hver kohort t, t=1,2,... Dersom en studerer tidsrekken av estimater for $p_t(x)$ for fast x og løpende t finner en at det er en betydelig variasjon fra kohort til kohort samtidig som det tydeligvis er sterk korrelasjon mellom ratene for ulike t. Vi betrakter følgelig settet av sannsynligheter $\underset{\sim}{p}_t = (\underset{\sim}{p}_t(x), x=0,16,..)$ som stokastiske. Disse varierer tilfeldig fra kohort til kohort omkring en eller annen trend. Kohortfordelingen $\underset{\sim}{p}_t$ antar vi har sannsynlighetstetthet $f_t(\underset{\sim}{p})$. I denne modellen har vi altså to typer usikkerhet. For gitte sannsynligheter har vi usikkerhet som avhenger av størrelsen på observasjonsmaterialet. Her gjelder sammenhengene

(1) $$E\left[\hat{p}_t(x) \,\middle|\, p_t(x)\right] = p_t(x),$$

(2) $$\mathrm{Var}\left[\hat{p}_t(x) \,\middle|\, p_t(x)\right] = p_t(x)(1-p_t(x))/N_t$$

og

(3) $$\mathrm{kov}\left[\hat{p}_t(x),\ \hat{p}_t(y) \,\middle|\, p_t(x),\ p_t(y)\right] = -p_t(x)p_t(y)/N_t$$

hvor $\hat{p}_t(x) = Y_t(x)/N_t$. Når N_t er stor ser vi at denne usikkerheten er liten. I tillegg har vi altså usikkerhet som skyldes variasjoner i de underliggende sannsynlighetene fra kohort til kohort.

I neste avsnitt beskrives en mulig fordelingsklasse for $\underset{\sim}{p}$ som har ønskelige matematiske egenskaper.

3. Multivariat betafordelte giftemålssannsynligheter.

Den multivariate betafordelingen (også kalt matrise-betafordelingen og Dirichletfordelingen) har tetthet

$$(4) \qquad f\,(p_1,p_2,\ldots p_k) = \frac{\Gamma(a_1+a_2+\ldots+a_k)}{\Gamma(a_1)\Gamma(a_2)\ldots\Gamma(a_k)}\, p_1^{\,a_1-1} p_2^{\,a_2-1}\ldots p_k^{\,a_k-1}$$

hvor $\Sigma p_j = 1$ og $\{a_j\}$ er parametre. Fordelingen har følgende egenskaper som er analoge til den multinomiske fordeling

$$(5) \quad Ep_j = a_j/a, \qquad a=\Sigma a_j,$$

$$(6) \quad \mathrm{Var}\ p_j = Ep_j\ (1-Ep_j)/(a+1)$$

og

$$(7) \quad \mathrm{Kov}(p_i,p_j) = -Ep_i\ Ep_j/(a+1)$$

Fordelingen er altså sentrert om et punkt som er bestemt ved de relative a-verdiene mens graden av konsentrasjon omkring dette punktet bestemmes av a. Mosiman (1962) studerer denne fordelingsklassen og viser at $\underset{\sim}{p}$ er multivariat betafordelt hvis og bare hvis p_j er "uavhengige på bibetingelsen $\sum_{j=1}^{k-1} p_j < 1$ nær". Hva dette betyr mer presist går vi ikke inn på her, intereserte henvises til Mosiman side 76.

I det følgende antar vi at $f_t\ (\underset{\sim}{p})$ er en multivariat beta-tetthet med parametre $a_t(x)$, $x=0,16,\ldots$ Betegner vi den multinomiske fordeling med parameter $\underset{\sim}{p}_t$ for $g_t(\underset{\sim}{y}_t,\underset{\sim}{p}_t)$ ($\underset{\sim}{p}_t$ gitt) får vi den ubetingede fordeling for $\underset{\sim}{y}_t$ ved

$$h_t\ (\underset{\sim}{y}_t) = \int_C g_t(\underset{\sim}{y}_t,\underset{\sim}{p}) f_t(\underset{\sim}{p})\, d\underset{\sim}{p} \qquad \text{(multippel integrasjon)}$$

hvor $C = \left\{ p(x) \,\middle|\, 0 \le p(x) \le 1,\ \sum_{x\ge 16} p(x) \le 1. \right\}$

Dette gir den multivariate negative hypergeometriske fordelingen (Jfr. Mosiman) gitt ved

$$(8) \qquad h_t(\underset{\sim}{y}_t) = \frac{N_t!\,\Gamma(a_t)}{\Gamma(a_t+N_t)}\ \prod_x \frac{\Gamma\ (a_t(x)+y_t(x))}{y_t(x)!\,\Gamma(a_t(x))}$$

Ved først å betinge med hensyn på $\underset{\sim}{p}_t$ kan vi beregne kovariansmatrisen til $\hat{\underset{\sim}{p}}_t(x)$. Fra (1) får vi

(9) $E\ \hat{p}_t(x) = E\ p_t(x) = a_t(x)/a_t$,

Bruker vi sammenhengen

$$\text{Var}\ \hat{p}_t(x) = E\ \text{Var}\ (\hat{p}_t(x) \mid \underset{\sim}{p}_t(x)) + \text{Var}\ E(\hat{p}_t(x) \mid p_t(x))$$

får vi

(10) $$\text{Var}\ \hat{p}_t(x) = \frac{(N_t+a_t)\ \text{Var}p_t(x)}{N_t}$$

Tilsvarende blir

(11) $$\text{cov}\left\{\hat{p}_t(x),\ \hat{p}_t(y)\right\} = \frac{(N_t+a_t)}{N_t}\ \text{kov}\left\{p_t(x),\ p_t(y)\right\}.$$

Vi merker oss at når N_t øker vil disse uttrykkene nærme seg (6) og (7).

Det kan vises at $Y_t(x)$ gitt $Y_t(j)$, $j \leq y < x$ er negativt hypergeometrisk fordelt med parametre $a_t(x)$ og $A_t(y)-a_t(x)$ hvor

$$A_t(y) = \sum_{j>y} a_t(j)$$

Dette medfører at

(12) $$E\ \{Y_t(x) \mid Y_t(j) = y_t(j),\ j \leq z < x\} = \frac{a_t(x)\,(N_t - \sum_{j \leq z} y_t(j))}{A_t(z)}$$

Det kan vises mer presist at når observasjonsmaterialet er stort er fordelingen (8) tilnærmet lik fordelingen (4). Mosiman (1962) gir en metode til å estimere parametrene $a_t(x)$ ved å bruke relasjonene (10) og (11) sammen med (6) og (7). En kan naturligvis også estimere parametrene ved å benytte sannsynlighetsmaksimeringsprinsippet med (8) som basis for "Likelihooden".

4. Teori for parametrene

Hittil har vi bare betraktet en kohort og ikke tatt hensyn til at det er en vesentlig sammenheng mellom kohortene. La $M_\tau(x)$ være antall ugifte menn ved tidspunkt τ i alder x og la $F_\tau(x)$ være antall ugifte kvinner ved tidspunkt τ i alder x. De $M_\tau(r)$, r=1,2,... ugifte mennene vil konkurrere med hverandre om å bli gift med de $F_\tau(r)$ r=1,2,.. kvinnene. Altså bør sannsynligheten for at en kvinne skal gifte seg i år τ gitt de tidligere giftemål i alle kohortene være en stigende funksjon i $M_\tau(r)$ og en avtagende funksjon i $F_\tau(r)$ for alle r. Denne endringen bør imidlertid være svakere jo større forskjellen er mellom x og den "ideelle" alder på partneren.

La

$$\mu_\tau(x) = E\{Y_{\tau-x}(x)/F_\tau(x) \mid \underset{\sim}{M}_{\tau-1}, \underset{\sim}{F}_{\tau-1}\}$$

hvor $\underset{\sim}{M}_\tau$ og $\underset{\sim}{F}_\tau$ er vektorer der komponentene er antatt ugifte i årene før τ i alle kohorter. Vi tenker oss nå at parametrene i sannsynlighetsfordelingen til $\underset{\sim}{p}$ for kohort t er funksjoner av typen $a_{t+x}(x) = a_\tau(x)$. Det betyr ifølge (12) at

$$(13) \qquad \mu_\tau(x) = a_\tau(x)/A\ (x)$$

fordi $\quad N_t - \sum_{j<x} y_t(j) = F_{t+x}(x)$

hvor $\quad A_\tau(x) = \sum_{j>0} a_{\tau+j}(x+j) = a - \sum_{j\geq 0} a_{\tau-j}(x-j)$

og a er en parameter, $a = \sum_x a_{t+x}(x)$.

Herav finner vi

$$a_\tau(16) = a\ \mu_\tau(16)$$

og

$$a_\tau(x) = \frac{\mu_\tau(x)\ a_{\tau-1}(x-1)}{(1+\mu_\tau(x))\mu_{\tau-1}(x-1)} = a \prod_{r=0}^{x-16} \frac{1}{1+\mu_{\tau-r}(x-r)} .$$

Det er altså en entydig sammenheng mellom $\{\mu_\tau(x)\}$ og $\{a_\tau(x)\}$. Neste skritt er nå å spesifisere $\mu_\tau(x)$ som funksjon av bestandsvektorene $\underset{\sim}{M}_\tau$ og $\underset{\sim}{F}_\tau$. La $f_\tau(x,y)$ være sannsynligheten for at en kvinne i alder x skal gifte seg med en mann i alder y i år τ. Problemet med å spesifisere f_τ er behandlet i litteraturen av flere forfattere. McFarland (1972) foreslår et sett aksiomer som f bør tilfredsstille hvorav de viktigste er

Aksiom 5. f(i,j) skal være en ikke-avtagende funksjon av M(i) når alle andre argumenter holdes faste. Videre skal det finnes et intervall hvor f(i,j) er strengt voksende i M(i) for et hvert fast sett av de andre argumentene. Det samme skal gjelde når kjønnene byttes om.

Aksiom 6. f(i,j) skal være en ikke-voksende (og på et eller annet intervall strengt avtagende) funksjon av M(k) for $k \neq i$ når alle andre argumenter holdes faste. Det samme skal gjelde når kjønnene byttes om. (Konkurranseaksiomet).

Aksiom 7. Den negative effekt på f(i,j) av en økning i M(r) skal være større enn den negative effekt av en ekvivalent økning i M(s) hvis r er nærmere i enn s er. Det samme skal gjelde når kjønnene byttes om. (Substitusjonsaksiomet).

Disse aksiomene betrakter McFarland som "minimumskrav" en bør stille til kandidater for giftemålsfunksjonen. Imidlertid er disse aksiomene restriktive nok til å ekskludere de fleste giftemålsfunksjoner som foreslås i litteraturen. Eksempler på slike funksjoner er gitt i McFarland (1972) og Pollard (1977). Pollard (1975) foreslår en funksjon av typen

$$(14) \qquad k_{xy}\, M_\tau(y) \Big/ \Big[\sum_j (\alpha_{xj} M_\tau(j) + \beta_{jy}\, F_\tau(j))\Big] .$$

hvor α_{ij} og β_{ji} er tiltrekningsparametre dvs. α_{ij} er et mål for tiltrekningen en kvinne i alder i har på en mann i alder j og β_{jy} er tiltrekningen en mann i alder y har på en kvinne i alder j. Denne funksjonsformen tilfredsstiller alle aksiomene så nær som aksiom 7. Pollard tror imidlertid at aksiom 7 er mindre vesentlig. Den overveiende delen av befolkningen søker en partner i nærheten av sin egen statusgruppe (aldersgruppe) og de velger en partner fra andre grupper bare dersom det er underskudd på kandidater i den ønskede gruppe. Denne observasjonen tyder på at de viktige her er "avstanden" mellom statusgruppene til henholdsvis menn og kvinner og at giftemålsfunksjonen bør være en avtagende funksjon i $|i-j|$. Schoen (1977) foreslår en giftemålsfunksjon av typen ovenfor hvor $\alpha_{xy} = \beta_{xy}$ og $k_{xy} = \alpha_{xy}(\alpha_{x\cdot} + \alpha_{\cdot y}) = \sum_j (\alpha_{xj} + \alpha_{jx})\, \alpha_{xy}$.

Hans modell begrunnes ved å innføre en rektangulær befolkning, dvs. en befolkning med likt antall personer (ugifte personer) i hver aldersgruppe og deretter innføre en korreksjonsfaktor som reflekterer endringer i kjønns og aldersammensetningen i den ugifte befolkningen. Vi skal i det følgende basere oss på denne funksjonen;

$$(15) \qquad W_{\tau}^{f}(x,y) = \frac{\alpha_{xy}(\alpha_{x\cdot} + \alpha_{\cdot y})\, M_{\tau}(y)}{\sum_{j} \alpha_{xj} M_{\tau}(j) + \sum_{j} \alpha_{jy} F_{\tau}(j)} .$$

hvor $W_{\tau}^{t}(x,y)$ er sannsynligheten for at en kvinne i alder x skal gifte seg med en mann i alder y ved tidspunkt τ.

Herav vår vi

$$\mu_{\tau}(x) = \sum_{y} W_{\tau}^{f}(x,y) \text{ og } \lambda_{\tau}(y) = \sum_{x} W_{\tau}^{m}(x,y)$$

hvor $\lambda_{\tau}(y)$ er de alderspesifikke giftemålssannsynlighetene for menn. Schoen har estimert parametrene α_{xy} på svenske data fra 1973. I vår situasjon hvor vi ikke har observasjoner av (x,y) - giftemål før 1970 må vi innskrenke parameter-rommet for å få effisient estimering. La oss se på hvilke egenskaper en slik glattingsfunksjon bør ha. α_{xy} bør være symmetrisk i x og $y-\nu(y)$ hvor $\nu(y)$ er en størrelse som tar hensyn til at menn ofte gifter seg med kvinner som er yngre enn dem selv. For y fast bør for α_{xy} ha toppunkt for $x = y-\nu(y)$. Funksjonen α_{xy} bør videre vokse raskt mot sin største verdi og deretter avta langsomt. Betrakt funksjonen

$$h(x,y) = \exp\{-\Theta_1 - \Theta_2 \,|x-y+\nu(y)| - \Theta_3(x-\nu(y)+y) + \Theta_4 \log (x-\nu(y)+y)\}$$

Partiell derivering m.h.p. x gir

$$\frac{\partial h}{\partial x} = h\left(\Theta_2 - \Theta_3 + \frac{\Theta_4}{x+y-\nu(y)}\right) \text{ for } x<y-\nu(y),$$

$$\frac{\partial h}{\partial x} = -h\left(\Theta_2 + \Theta_3 - \frac{\Theta_4}{x+y-\nu(y)}\right) \quad \text{for } x>y-\nu(y).$$

Dersom $\Theta_2 > \Theta_3 - \Theta_4/(32-\nu(16))$

vil

$$\frac{\partial h}{\partial x} > 0 \text{ for } x<y-\nu(y)$$

og

$$\frac{\partial h}{\partial x} < 0 \text{ for } x > y - \nu(y)$$

hvilket betyr at h har sin største verdi for $x=y-\nu(y)$. La $x_1<y-\nu(y)$ og $x_2 > y-\nu(y)$ være slik at $h(x_1,y) = h(x_2,y)$. Vi har at

$$\left|\frac{\partial h}{\partial x}(x_1,y)\right| - \left|\frac{\partial h}{\partial x}(x_2,y)\right| = h(x_1,y)\left(\frac{\Theta_4}{x_1+y-\nu(y)} + \frac{\Theta_4}{x_2+y-\nu(y)} - 2\Theta_3\right).$$

Dersom nå $\Theta_4 > \Theta_3(2M-\nu(M))$ hvor M er høyeste giftemålsalder vil differansen ovenfor være positiv. Dette betyr at endringen i h skjer raskere når $x<y-\nu(y)$ enn når $x>y-\nu(y)$. Vi kan videre generalisere h ved å innføre høyere ordens ledd i nevneren.

Funksjonen $\nu(y)$ kan f.eks. velges lineær; $\nu(y) = ry+s$ hvilket gir $y-\nu(y) = (1-r)y-s = r_1y-s$, der $r_1=1-r$. Her kan vi <u>ikke</u> la r og s være parametre som estimeres simultant med de andre parametre. Grunnen til dette er at de partiellderiverte m.h.p. r og s ikke eksisterer i punktet $x=y-\nu(y)$ fordi funksjonen h inneholder et tallverdiledd i eksponenten. De vanlige programmer for iterasjon forutsetter nemlig eksistensen av de partiellderiverte m.h.p. parametrene. En måte å teste funksjonsformen på er å estimere parametrene α_{ij} på grunnlag av (i,j)-datamatrisene etter 1970 og deretter glatte disse estimatene. Disse estimatene har jo dessuten interesse i seg selv. Vi kan nå estimere $W_\tau(x,y)$ for hvert tidspunkt med utgangspunkt i de observerte ratene. Vi vil da sannsynligvis finne at parametrene varierer med tiden. Det kan derfor være aktuelt å anta at periodefordelingen følger en lineær stokastisk modell f.eks. en førsteordens autoregressiv modell

(16) $$q_{\tau+1}(x) - \mu_{\tau+1}(x) = \phi_1(q_\tau(x) - \mu_\tau(x)) + \phi_2(q_\tau(x-1) - \mu_\tau(x-1)) + \varepsilon_{\tau+1}(x)$$

hvor $q_\tau(x)$ er sannsynligheten for at en ugifte kvinne i alder x gifter seg ved tidspunkt τ, $\varepsilon_\tau(x)$ er uavhengige for alle τ og x og $\varepsilon_{\tau+1}(x)$ er uavhengig av $q_\tau = \{q_\tau(x), x = 16,17,..\}$.

ϕ_1 og ϕ_2 er ukjente parametre som har tallverdi mindre enn 1. Aktuelle hypoteser er her at enten ϕ_1 eller ϕ_2 er lik null, eventuelt at begge to er lik null. Dersom begge parametrene er lik null er vi tilbake til situasjonen som ble beskrevet i begynnelsen av avsnittet. Da er nemlig den vesentlige variasjonen i giftemålssannsynlighetene forklart ved (15). Dersom ϕ_1 og ϕ_2 er forskjellig fra null betyr det at q_τ i lengre eller kortere perioder aviker fra μ_τ. Siden

(17) $$E\{q_\tau(x) | M_\tau, F_\tau\} \doteq \frac{a_\tau(x) + Y_{\tau-x}(x)}{A_\tau(x) + F_\tau(x)}$$

får vi at

$$(18)\qquad a_{\tau+1}(x)/A_{\tau+1}-\mu_{\tau+1}(x) = \sum_{j=0}^{1} \Phi_{j+1}\{\frac{a_{\tau}(x-j)+Y_{\tau-x+j}(x-j)}{A_{\tau}(x-j) + F_{\tau}(x-j)} - \mu_{\tau}(x-j)\}$$

Herav kan vi beregne $a_{\tau} = \{a_{\tau}(x), x=16,17,..\}$ rekursivt.

La nå

$$G^{f}(y_{t}(x)) = \Pr\{Y^{f}_{\tau-x}(x) = y_{\tau}(x) | \underset{\sim}{M}_{\tau-1}, \underset{\sim}{F}_{\tau-1}\}$$

Det følger fra avsnitt (3) at $Y_{\tau}(x)$ er negativt hypergeometrisk fordelt med parametre $A_{\tau}(x) - a_{t}(x)$ og $a_{\tau}(x)$ dvs.

$$(19)\qquad G^{f}_{\tau}(y_{\tau}(x)) = \frac{F_{\tau}(x)!\Gamma(A_{\tau}(x))\Gamma(a_{\tau}(x) + y_{\tau}(x))\Gamma(A_{\tau}(x) + F_{\tau}(x)-a_{\tau}(x)-y_{\tau}(x))}{\Gamma(A_{\tau}(x) + F_{\tau}(x))\; y_{\tau}(x)!(F_{\tau}(x)-y_{\tau}(x))!\Gamma(A_{\tau}(x)-a_{\tau}(x))\Gamma(a_{\tau}(x))}.$$

Hittil har vi bare sett på giftemålssannsynligheter for kvinner. Med de samme forutsetninger for giftemålsannsynligheten for menn som for kvinner får vi helt analoge resultater for menn som ovenfor.

5. Estimering

Vi skal se på estimering ved bruk av minste kvadraters metode og ved sannsynlighetsmaksimeringsprinsippet. Et problem her er at antall kvinner som gifter seg er lik antall menn som gifter seg ved hvert tidspunkt slik at en dermed får stokastisk avhengighet som skyldes denne typen kjønnsavhengighet. Hoem (1969) viser hvilke matematiske problemer som oppstår dersom en skal ta hensyn til denne avhengigheten under estimeringen. Dersom vi imidlertid estimerer modellene for hvert kjønn separat trenger vi ikke å spesifisere den stokastiske avhengigheten mellom hannkjønns- og hunkjønnsratene. I vår situasjon hvor antall observasjoner for hvert kjønn er stort (minst 2 600) er dette en aktuell mulighet. Vi antar at (16) gjelder også når sannsynligheten erstattes med de observerte ratene $\hat{q}_{\tau}(x)$. Siden restleddene $\varepsilon_{\tau}(x)$ er antatt å være stokastisk uavhengige for alle τ og alle x får vi at

$$U_{\tau}(x) = \hat{q}_{\tau+1}(x)-\mu_{\tau+1}(x) - \sum_{j=0}^{1} \phi_{1+j}(\hat{q}_{\tau}(x-j)-\mu_{\tau}(x-j))$$

for alle τ og x er uavhengige. Minste kvadraters estimator finnes derfor ved å minimere

$$(20) \qquad \sum_{\tau,x} U_\tau(x)^2$$

med hensyn på parametrene ϕ_j og parametrene Θ_j i glattingsfunksjonen for $\{\alpha_{xy}\}$. Tilsvarende kan vi benytte giftemålsratene for menn til å estimere ψ_j og θ_j hvor ψ_j er de parametrene som tilsvarer ϕ_j i (16).

Vi skal dernest betrakte estimering ved hjelp av sannsynlighetsmaksimeringsprinsippet. Dette gir oss mulighet til å ta hensyn til forutsetningene i avsnitt 2 under estimeringen. Benytter vi bare giftemålsratene for kvinner, får vi at likelihooden for observasjonene blir lik

$$(21) \qquad L = \prod_\tau \prod_{r,x} G_\tau^f(y_\tau(x)).$$

Dersom observasjonsmaterialet er stort er fordelingen til $\{\hat{q}_\tau(x)\}$ tilnærmet lik fordelingen til $\{q_\tau(x)\}$. Følgelig har vi da

$$(22) \qquad G_\tau^f\,(y_t(x)) \simeq \frac{\Gamma(A_\tau(x))\;\hat{q}_\tau(x)^{a_\tau(x)-1}\;(1-\hat{q}_\tau(x))^{A_\tau(x)-a_\tau(x)-1}}{\Gamma(a_\tau(x))\Gamma(A_\tau(x)-a_\tau(x))}$$

og tilsvarende uttrykk for G_τ^m.

Antar vi videre at

$$\frac{a_\tau(x) + Y_{\tau-x}(x)}{A_\tau(x) + F_\tau(x)} \simeq Y_{\tau-x}(x)/F_\tau(x) = \hat{q}_\tau(x)$$

får vi ved å benytte resonementet på side 5 at

$$(23) \qquad a_\tau(x) = a\prod_{j=0}^{x-16} \frac{1}{[\mu_{\tau-j}(x-j)+\phi_1(\hat{q}_{\tau-1-j}(x-j)-\mu_{\tau-1-j}(x-j)) + \phi_2(\hat{q}_{\tau-1-j}(x-1-j)-\mu_{\tau-1-j}(x-1-j))]}.$$

For parametrene knyttet til giftemålssannsynlighetene for menn får vi analoge uttrykk.

REFERANSER

Hoem, J.M., 1969: Concepts of a Bisexual Theory of Marriage Formation. Statistisk Tidskrift **4**, 295-300

McFarland, D.D., 1972: Comparison of alternative marriage models. Populations Dynamics (Greville, T., Ed), 89-106. Academic Press, New York.

Mosiman, J.E., 1962: On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. Biometrica, **49**, 65-82.

Pollard, J.H., 1975: Modelling human populations for projection purposes - some of the problems and challenges. Aust. J. Statist. **17**, 63-76.

Pollard, J.H., 1977: The continuing attempt to incorporate both sexes into marriage analysis. International population conference **1**, 291-308. Mexico 1977.

Schoen, R., 1977: A two-sex nuptiality-mortality life table. Demography **14**, 333-347.