Arbeidsnotater

STATISTISK SENTRALBYRÅ

IO 68/11 Oslo, 5. juni 1968

Befolkningsprognosemodellens flyttingsrelasjoner. I.

Av Jan M. Hoem.

Innhold

Forord

Dette notatet representerer vårt første forsøk på å angripe problematikken rundt de interne flyttingene her i landet. Hovedhensikten denne gang er å vurdere hvilken matematisk apparatur det kan lønne seg å operere med. Jeg vil gjerne understreke at notatet ikke representerer mer enn en løs idéskisse.

1. Generelle innledningsbetraktninger

§ 1A. Når vi skal studere innenlandske flyttinger i Norge, kan det være fruktbart å tenke seg landet delt opp i N geografisk bestemte områder, som vi nummererer fra 1 til N. La oss betrakte to slike områder, nr. i og nr. j. Sjansene for at en person fra i-te område skal flytte til j-te område i løpet av et visst tidsrom, avhenger av visse hovedfaktorer som vi kan gruppere slik:

(i) Egenskaper ved i-te område.

(ii) Egenskaper ved j-te område.

(iii) Egenskaper ved de N-2 andre områdene i landet, idet disse "konkurrerer" med nr. i og nr. j om å være oppholdssted for den personen vi betrakter.

(iv) Karakteristika ved personen selv.

Ved siden av disse faktorene må vi ta i betraktning at personen kan "forsvinne ut av systemet" ved død eller utvandring. Denne muligheten virker i grunnen som et forstyrrende element i våre undersøkelser, men vi må passe på å ta hensyn til den der det er nødvendig.

Vi skal se hvordan man etter hvert kan trekke inn i analysen de hovedfaktorene vi har regnet opp, og begynner med den siste gruppen.

§ 1B. (Personlige karakteristika.) Det er mange statistisk kvantifiserbare personlige karakteristika som kan ha innflytelse på flyttetilbøyeligheten. Viktige eksempler på dette er alder, kjønn, yrke, utdannelse, familieforhold, og tidligere flyttingshistorie.

På det nåværende stadium skal vi regne med at det er områdenes og personenes aktuelle karakteristika som bestemmer flyttetilbøyelighetene. Vi skal ikke eksplisitt trekke inn mulige virkninger av hvordan disse egenskapene er ervervet. Visse teorier går ut på at en person vil ha en stadig synkende tendens til å flytte fra et sted jo lenger han har bodd der ("cumulative inertia"). Vi vil ikke ta med dette momentet.

Det er mulig at vi kan komme tilbake til slikt senere i arbeidet med flytteproblematikken.

Av de personlige karakteristika står personens alder i en særstilling. Vi vil sørge for å ta med denne sentrale variable overalt i analysen.

De øvrige karakteristika kan deles inn i to typer: (a) medfødte egenskaper, og (b) ervervede karakteristika. La oss se litt på disse to kategoriene hver for seg.

§ 1C. Av medfødte egenskaper skal vi bare trekke inn personens kjønn. Siden flyttetilbøyeligheten er forskjellig for menn og kvinner, er det naturlig å utføre en separat analyse for hvert kjønn. I det følgende vil dette være underforstått. Vi kommer derfor ikke til å ta med kjønnsangivelse på de symboler vi innfører.

§ 1D. Alle de øvrige eksemplene vi gav i § 1B, er ervervede karakteristika. Disse kan endre seg etter som tiden går, og slike forandringer medfører endringer i flyttetilbøyeligheten. Omvendt vil en persons flyttehistorie opplagt ha tilbakevirkning på de egenskapene som erverves. (Man kan flytte for å få (mer) utdannelse, for å komme til et bedre ekteskapsmarked, for å gå inn i et nytt yrke, osv.) Det er altså et gjensidig samvirkningsforhold mellom en persons ervervede karakteristika og hans flyttetilbøyelighet. Dette leder til at vi ønsker å utføre en simultan analyse av flyttinger og ervervede karakteristika. Vi skal bruke kapittel 2 til å antyde hvordan det kan gjøres.

2. Simultan analyse av flyttinger og ekteskapsstatus.

§ 2A. Som en illustrasjon kan vi la en persons ekteskapsstatus være representant for de ervervede karakteristika, slik at vi ikke trekker andre av dem inn i analysen. Hvis man vil ha noe å knytte tanken til, kan man jo operere med statusene ugift, gift, enke/enkemann, separert, og skilt. Fremstillingen er imidlertid ikke bundet til dette settet.

§ 2B. La $P_{(i,m),(j,n)}(x,t)$ være sannsynligheten for at en x år gammel person som bor i i-te område og har ekteskapsstatus m, skal bo i j-te område og ha ekteskapsstatus n når han blir x+t år gammel. Her kan vi godt ha i=j og/eller m=n. Vi definerer

$$(2.1)\quad \phi_{ijm}(x) = \lim_{t\downarrow 0} P_{(i,m),(j,m)}(x,t)/t \quad \text{for } i\neq j,$$

og kaller dette for flytteintensiteten fra område i til område j for ekteskapsstatus m og alder x. Denne størrelsen tolkes slik at $\phi_{ijm}(x)\,dx$ er sannsynligheten for at en x år gammel person som bor i i-te område og har ekteskapsstatus m, skal flytte til j-te område innen alder x+dx (uten også å skifte ekteskapsstatus, dø, eller emigrere). Det er $\phi_{ijm}(x)$ som er vårt mål på det vi løselig har kalt flyttetilbøyeligheten (for en x-åring i i-te område og med ekteskapsstatus m).

Analogt definerer vi

(2.2) $$\nu_{imn}(x) = \lim_{t\downarrow 0} P_{(i,m),(i,n)}(x,t)/t \quad \text{for } m\neq n.$$

Hvis vi regner med de fem mulighetene for ekteskapsstatus som ble nevnt i § 2A, kan vi kalle $\nu_{i12}(x)$ for giftermålsintensiteten (for område i og alder x), $\nu_{i25}(x)$ er skilsmisseintensiteten, osv. Vi vil få $\nu_{i34}(x) = \nu_{i35}(x) = \nu_{i53}(x) = \nu_{i54}(x) = 0$ og $\nu_{im1}(x) = 0$ for m>1.

Vi vil ikke tillate at noen skal kunne flytte og skifte ekteskapsstatus akkurat samtidig, så vi vil forlange at

(2.3) $$\lim_{t\downarrow 0} P_{(i,m),(j,n)}(x,t)/t = 0 \quad \text{når både } i\neq j \text{ og } m\neq n.$$

§ 2C. Vi innfører også sannsynligheten $P_{(i,m),(N+1,n)}(x,t)$ for at en x år gammel person som bor i i-te område og har ekteskapsstatus m, skal utvandre fra landet innen alder x+t og ha ekteskapsstatus n ved emigrasjonen. Emigrasjonsintensiteten blir

(2.4) $$\eta_{im}(x) = \lim_{t\downarrow 0} P_{(i,m),(N+1,m)}(x,t)/t \quad \text{for } i\leq N,$$

og vi krever at

(2.5) $$\lim_{t\downarrow 0} P_{(i,m),(N+1,n)}(x,t)/t = 0 \quad \text{for } i\leq N,\ m\neq n,$$

idet vi ikke vil tillate at emigrasjon og skifte av ekteskapsstatus faller sammen i tid.

§ 2D. La oss her se bort fra muligheten av re-immigrasjon.

§ 2E. Endelig skal vi innføre sannsynligheten $P_{(i,m),(N+2,n)}(x,t)$ for at en x år gammel person som bor i i-te område og har ekteskapsstatus m, skal dø innen alder x+t og ha ekteskapsstatus n ved dødsfallet. Dødsintensiteten blir

$$(2.6) \quad \mu_{im}(x) = \lim_{t \downarrow 0} P_{(i,m),(N+2,m)}(x,t)/t \quad \text{for } i \leq N,$$

og vi vil kreve at

$$(2.7) \quad \lim_{t \downarrow 0} P_{(i,m),(N+2,n)}(x,t)/t = 0 \quad \text{for } i \leq N,\ m \neq n.$$

§ 2F. Hvis vi kjenner alle $\phi_{ijm}(x)$, $\nu_{imn}(x)$, $\eta_{im}(x)$, og $\mu_{im}(x)$, er det prinsipielt mulig å finne alle ønskelige verdier av alle $P_{(i,m),(j,n)}(x,t)$. Som de ovenstående paragrafene antyder, vil en simultan analyse av ekteskapsstatus og interne flyttinger utføres ved at man samtidig trekker inn emigrasjon og dødelighet. Analysen består da i et studium av overgangsintensitetene ϕ, ν, η, og μ.

Vi skal ikke her ta opp estimeringsteorien for denne modellen. Vårt hovedformål er nå bare å innføre de grunnleggende begrepene man har å arbeide med.

§ 2G. Vi har her sett hvordan det er mulig å trekke inn en persons ekteskapsstatus i en analyse av flyttingene. Andre ervervede karakteristika kan behandles analogt. I de følgende kapitler skal vi imidlertid gå den motsatte vei og utelate alle ervervede karakteristika. Dette vil forenkle fremstillingen noe, og det tillater oss bedre å konsentrere oss om vårt hovedtema, flyttingene. Vi får så føre inn relevante ervervede karakteristika når det blir nødvendig.

§ 2H. I det opplegg vi her har fulgt, er det den enkelte person vi betrakter som den flyttende enhet. Det er kanskje grunn til å overveie om det ikke ville være fordelaktig å bruke familien som enhet istedenfor. Særlig gjelder det vel for små barn at de flytter sammen med familien, fordi familien forøvrig flytter. Indirekte kommer dette momentet inn ved at vi betrakter individets familieforhold som en faktor som kan påvirke flyttetilbøyeligheten. Det er allikevel et langt skritt fra dette og til å betrakte familien som flyttende enhet.

Foreløpig synes det imidlertid å være alt for mange problemer knyttet til en slik fremgangsmåte, så vi bør vel fortsette å betrakte individet som flytte-enhet i alle fall inntil videre.

3. Generasjonsavhengige funksjoner.

§ 3A. Erfaring viser at flytteintensiteter og andre "demografiske intensiteter" av den typen vi definerte i forrige kapittel, vil variere fra et fødselskull til et annet. Således må man f.eks. operere med ēn funksjon $\phi_{ijm}(x)$ for personer født i perioden 1931-35, og en annen slik funksjon for personer født i perioden 1941-45, osv. Dette kunne vi angitt ved å innføre fødselsdato som funksjonsargument, slik at $\phi_{ijm}(x,\tau)$ ble flytteintensiteten for personer født på tidspunkt τ.

En slik angivelse vil bli brukt i det følgende.

§ 3B. La nå $P_{ij}(x,t,\tau)$ være sannsynligheten for at en person som er født på tidspunkt τ og som x år gammel bor i i-te område, skal (være i live og) bo i j-te område i alder $x+t$. Her er i og $j \leqq N$. La videre $P_{i,N+1}(x,t,\tau)$ være sannsynligheten for at han skal utvandre innen alder $x+t$, og la $P_{i,N+2}(x,t,\tau)$ være sannsynligheten for at han skal dø i Norge innen alder $x+t$. Vi innfører flytteintensiteten

$$(3.1) \qquad \phi_{ij}(x,\tau) = \lim_{t \downarrow 0} P_{ij}(x,t,\tau)/t \qquad \text{for } i \neq j, \text{ i og } j \leqq N,$$

og definerer en emigrasjonsintensitet $\eta_i(x,\tau)$ og en dødsintensitet $\mu_i(x,\tau)$. En (partiell) analyse av flyttingene består nå i et studium av disse intensitetene.

§ 3C. Som nevnt i § 1A kan man innta det standpunkt at emigrasjonen og dødsfall representerer "forstyrrelser" som man helst vil eliminere når man skal studere de interne flyttingene. En måte å gå frem på er da å utelate fra analysen opplysninger om dem som er emigrert eller falt fra i løpet av den observasjonsperiode man har, og kun basere seg på den resterende bestand. Det er viktig å merke seg at dette leder til en analyse av noe annet enn de funksjonene $\phi_{ij}(x,\tau)$ som vi definerte i (3.1).

Anta nemlig at man sørger for å få data for personer født på tidspunkt τ fra de er x til de er $x+\zeta$ år gamle, og at man "luker vekk" opplysningene vedrørende dem som er emigrert eller døde i perioden. Hvis vi lar $P^{*}_{ij}(y,t,\tau)$ (for $x \leq y \leq y+t \leq x+\zeta$) betegne sannsynligheten for at en av de gjenværende som y år gammel bor i i-te område, skal bo i j-te område i alder $y+t$, er det lett å innse at

$$(3.2) \quad P^{*}_{ij}(y,t,\tau) = \frac{P_{ij}(y,t,\tau) \sum_{k=1}^{N} P_{jk}(y+t,x+\zeta-y-t,\tau)}{\sum_{k=1}^{N} P_{ik}(y,x+\zeta-y,\tau)} .$$

Om vi dividerer med t og lar $t \downarrow 0$, ser vi at den tilsvarende intensiteten blir

$$(3.3) \quad \phi^{*}_{ij}(y,\tau) = \phi_{ij}(y,\tau) \cdot \frac{\sum_{k=1}^{N} P_{jk}(y,x+\zeta-y,\tau)}{\sum_{k=1}^{N} P_{ik}(y,x+\zeta-y,\tau)} \quad \text{for } i \neq j.$$

Siden denne brøken ikke generelt er lik 1, blir $\phi^{*}_{ij}(y,\tau) \neq \phi_{ij}(y,\tau)$. Man har dessuten bare tilsynelatende frigjort seg fra innflytelsen fra emigrasjon og dødelighet, for brøken i (3.3) vil avhenge av funksjonene $\eta_i(x,\tau)$ og $\mu_i(x,\tau)$.

Man bør derfor ikke "luke" observasjonsmaterialet før analysen utføres.

§ 3D. Overgangsintensiteter som dem vi definerte i § 3B vil endre seg med τ, men innenfor begrensede tidsrom vil variasjonene neppe være særlig store. Vi vil derfor regne med at vi kan finne tidsintervaller av typen $\langle\tau_0,\tau_1]$ der vi i hvert fall tilnærmet kan sette $\phi_{ij}(x,\tau) = \phi_{ij}(x,\tau')$ for alle $\tau \in \langle\tau_0,\tau_1]$, alle i, j, og x, der τ' er et passende, fast tidspunkt i $\langle\tau_0,\tau_1]$. Vi vil anta at tilsvarende relasjoner kan stilles opp for alle η_i og μ_i.

I dette notatet vil vi vanligvis betegne et intervall av typen $\langle\tau_0,\tau_1]$ som "periode n". Vi vil bruke symbolene $\phi_{ij}(x;n)$, $\eta_i(x;n)$ og $\mu_i(x;n)$ for intensitetene for fødselskullet fra denne perioden.

§ 3E. Vi skal nå ta opp litt av teorien for hvordan intensitetene kan estimeres hvis man har data for et gitt kull av personer født i en gitt periode n. La oss si at fullstendige data foreligger for disse personene for aldre mellom a og b. Vi deler inn intervallet $\langle a,b]$ i passende delintervaller $\langle a_0,a_1]$, $\langle a_1,a_2]$, ..., $\langle a_{K-1},a_K]$, med $a_0=a$ og $a_K=b$. Hvert intervall $\langle a_{k-1},a_k]$ velges slik at alle $\phi_{ij}(x;n)$, alle $\eta_i(x;n)$, og alle $\mu_i(x;n)$ varierer lite når x gjennomløper intervallet. For alle $x \in \langle a_{k-1}, a_k]$ erstatter vi funksjonsverdien $\phi_{ij}(x;n)$ med en konstant $\phi_{ij}'(k,n)$. Analogt innfører vi parametre $\eta_i'(k,n)$ og $\mu_i'(k,n)$ for funksjonsverdiene $\eta_i(x;n)$ og $\mu_i(x;n)$ når x ligger i dette intervallet.

La nå $F_{ij}(k,n)$ være antall flyttinger $i \to j$ observert for $x \in \langle a_{k-1}, a_k]$, la $U_i(k,n)$ være det observerte antall emigrasjoner fra i-te område i dette aldersintervallet, la $D_i(k,n)$ være det tilsvarende antall dødsfall, og la $B_i(k,n)$ være samlet observert levetid (målt i personår, dvs. "bestand under risiko"). Da estimeres parametrene ved

$$(3.4) \qquad \hat{\phi}_{ij}'(k,n) = F_{ij}(k,n)/B_i(k,n), \quad \hat{\eta}_i'(k,n) = U_i(k,n)/B_i(k,n),$$

$$\text{og} \quad \hat{\mu}_i'(k,n) = D_i(k,n)/B_i(k,n).$$

Vi bruker altså aldersårsmetoden.

§ 3F. Vi skal så se hvordan vi kan beregne sannsynlighetene $P_{ij}(x,t;n)$ for $i \leq N$, $j \leq N+2$ når intensitetsparametrene er kjent. Vi definerer

$$(3.5) \qquad \phi_i(x,\tau) = \sum_{j=1}^{N} \phi_{ij}(x,\tau),$$

og kaller dette for total <u>utflyttingsintensitet</u> for område i (og alder x for generasjonen født på tidspunkt τ). Videre definerer vi

$$(3.6) \qquad \rho_i(x,\tau) = \phi_i(x,\tau) + \eta_i(x,\tau) + \mu_i(x,\tau),$$

og kaller dette for den tilsvarende totale <u>avgangsintensitet</u>. Når vi går fram slik som i § 3D, vil vi erstatte alle funksjonsverdiene $\rho_i(x;n)$ for $x \in \langle a_{k-1}, a_k]$ med én parameterverdi $\rho_i'(k,n)$.

Vi innfører matrisen

$$Q(k,n) = \begin{bmatrix} -\rho_1', & \phi_{12}', \ldots, \phi_{1N}', & \eta_1', & \mu_1' \\ \phi_{21}', & -\rho_2', \ldots, \phi_{2N}', & \eta_2', & \mu_2' \\ \cdot & \cdot \quad \cdots \quad \cdot & \cdot & \cdot \\ \phi_{N1}', & \phi_{N2}', \ldots, -\rho_N', & \eta_N', & \mu_N' \\ 0, & 0, \ldots, 0, & 0, & 0 \\ 0, & 0, \ldots, 0, & 0, & 0 \end{bmatrix},$$

der alle $\phi_{ij}' = \phi_{ij}'(k,n)$ og tilsvarende gjelder for alle ρ_i', η_i', μ_i'. La $P_{ij}'(x,t,n)$ være den tilnærmelsesverdi til $P_{ij}(x,t;n)$ som vi får ved overalt å innføre parametrene $\phi'_{ij}(k,n)$ osv. for funksjonsverdiene $\phi_{ij}(x;n)$ osv. Vi innfører matrisen

$$P'(x,t,n) = \begin{bmatrix} P'_{11}, & P'_{12}, & \dots, & P'_{1N}, & P'_{1,N+1}, & P'_{1,N+2} \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdot \\ P'_{N1}, & P'_{N2}, & \dots, & P'_{NN}, & P'_{N,N+1}, & P'_{N,N+2} \\ 0, & 0, & \dots, & 0, & 1, & 0 \\ 0, & 0, & \dots, & 0, & 0, & 1 \end{bmatrix},$$

der $P'_{ij} = P'_{ij}(x,t,n)$ overalt. Da er

$$(3.7) \qquad P'(x,t,n) = \sum_{\nu=0}^{\infty} \frac{t^\nu}{\nu!} Q^\nu(k,n) \quad \text{for } a_{k-1} \leq x \leq x+t \leq a_k .$$

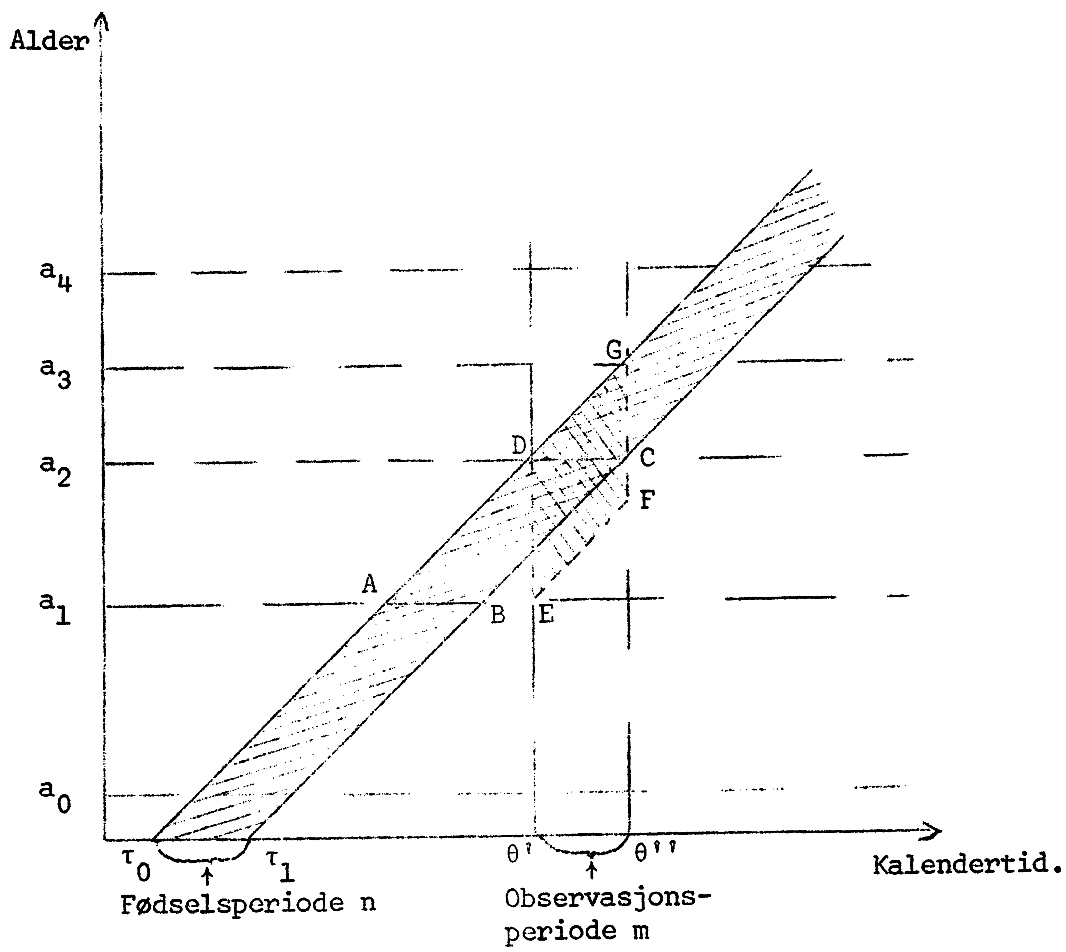
Hvis $\hat{Q}(k,n)$ er vår estimator for $Q(k,n)$ og $\hat{P}'(x,t,n)$ analogt er vår estimator for $P'(x,t,n)$, gjelder en relasjon helt analog til (3.7) for estimatormatrisene.

Hvis $x \in [a_{k-1}, a_k\rangle$ og $x+t \in \langle a_{m-1}, a_m]$ der $m>k$, får vi

$$(3.8) \qquad P'(x,t,n) = P'(x,a_k-x,n) \prod_{\nu=k+1}^{m-1} P'(a_{\nu-1},a_\nu-a_{\nu-1},n) \,.\, P'(a_{m-1},x+t-a_{m-1},n)$$

der $\prod_{\nu=k+1}^{m-1} P'(a_{\nu-1},a_\nu-a_{\nu-1},n)$ erstattes med 1 når $m=k+1$. Hver av faktorene på høyre side i (3.8) beregnes ved hjelp av (3.7).

§ 3G. Nedenstående figur representerer et utsnitt av Lexis skjema. For å kunne beregne estimatene i (3.4) for k = 1,2,...,K, trenger vi data fra det området som er høyreskravert på figuren.



4. Periodeavhengige funksjoner.

§ 4A. Av forskjellige viktige administrative grunner foreligger ikke data ordnet etter fødselskull, men etter kalenderperiode for observasjonstidspunktet. Vanligvis er dataene således ordnet etter observasjonsår.

Når man har så detaljerte data som i det norske personregisteret, er det mulig å omorganisere dataene slik at man kan utføre en generasjonsanalyse. Hvis en slik analyse skal strekke seg ut over noen få aldersår, blir det imidlertid nødvendig å operere med data fra mange kalenderår samtidig. Dette vil skape visse tekniske vanskeligheter som vel kan bli ganske store. Dertil kommer at vi ikke har tilgjengelig data fra så mange kalenderår ennå.

Med bakgrunn i overlegninger som disse, skal vi definere periodeavhengige funksjoner analoge til de generasjonsavhengige funksjonene vi opererte med i kapittel 3. De periodeavhengige funksjonene er også bedre tilpasset de behov man har ved beregning av befolkningsprojeksjoner.

§ 4B. La $\bar{P}_{ij}(x,t,\theta)$ for i og $j \leq N$ være sannsynligheten for at en person som på tidspunkt θ er x år gammel og bor i i-te område, skal bo i j-te område i alder x+t. $\bar{P}_{i,N+1}(x,t,\theta)$ og $\bar{P}_{i,N+2}(x,t,\theta)$ defineres analogt. Siden denne personen jo må ha $\tau = \theta - x$ til fødselstidspunkt, blir

$$\bar{P}_{ij}(x,t,\theta) = P_{ij}(x,t,\theta-x) \quad \text{for } i \leq N,\ j \leq N+2. \tag{4.1}$$

De tilsvarende intensiteter er

$$\begin{aligned} &\bar{\phi}_{ij}(x,\theta) = \phi_{ij}(x,\theta-x),\ \bar{\eta}_i(x,\theta) = \eta_i(x,\theta-x), \\ &\text{og } \bar{\mu}_i(x,\theta) = \mu_i(x,\theta-x). \end{aligned} \tag{4.2}$$

I kapittel 3 innførte vi så intensitetsparametre $\phi_{ij}^{!}(k,n)$ osv., definert for fødselsperioden n. Analogt skal vi nå innføre parametre $\bar{\phi}_{ij}^{!}(k,m)$, definert for <u>observasjons</u>-perioden m, som er tidsrommet $\langle\theta', \theta''\rbrack$, ved tilnærmelsen

$$\bar{\phi}_{ij}(x,\theta) \approx \bar{\phi}_{ij}^{!}(k,m) \quad \text{for alle } \theta \in \langle\theta',\theta''\rbrack,\ x \in \langle a_{k-1}+\theta-\theta'', a_k+\theta-\theta'\rbrack. \tag{4.3}$$

Det litt pussige variasjonsområdet for x her motiveres slik: En person som er x år gammel på tidspunkt θ, vil være $x-(\theta-\theta')$ år på tidspunkt θ'. Når x ligger i intervallet angitt i (4.3), vil $x-(\theta-\theta') \in \langle a_{k-1}, a_k\rbrack$. $\bar{\phi}_{ij}^{!}(k,m)$ blir da flytteintensitetparameteren i periode m for dem som hadde alder i intervallet $\langle a_{k-1}, a_k\rbrack$ ved begynnelsen av perioden.

Parametre $\bar{\eta}_i^{!}(k,m)$ og $\bar{\mu}_i^{!}(k,m)$ defineres analogt.

Slik vi har foretatt våre tilnærmelser, kan vi ikke regne med at det vil finnes en relasjon av typen (4.2) som forbinder f.eks. $\phi_{ij}{}'(k,n)$-ene med $\bar{\phi}_{ij}{}'(k,m)$-ene. Vi ser et eksempel på dette i Lexis-skjemaet i § 3G. Parameteren $\phi_{ij}^{!}(2,n)$ vil der referere seg til området ABCD, mens f.eks. $\bar{\phi}_{ij}^{!}(2,m)$ refererer seg til området EFGD. Disse parametrene er sannsynligvis forskjellige, selv om forskjellen kanskje ikke er stor.

§ 4C. I analogi med (3.4) gir kalenderårsmetoden estimatorer på formen

$$\begin{aligned} &\hat{\bar{\phi}}_{ij}^{!}(k,n) = \bar{F}_{ij}(k,m)/\bar{B}_i(k,m),\ \hat{\bar{\eta}}_i^{!}(k,m) = \bar{U}_i(k,m)/\bar{B}_i(k,m), \\ &\text{og } \hat{\bar{\mu}}_i^{!}(k,m) = \bar{D}_i(k,m)/\bar{B}_i(k,m). \end{aligned} \tag{4.4}$$

§ 4D. La nå $\bar{P}'_{ij}(x,t,m)$ være den tilnærmelsesverdi til $\bar{P}_{ij}(x,t,m)$ vi får ved overalt å innføre parametrene $\bar{\phi}'_{ij}(k,m)$ osv. for funksjonene $\bar{\phi}_{ij}(x,\theta)$ osv. Da er

$$(4.5)\qquad \bar{P}'(x,t,\tau) = \sum_{\nu=0}^{\infty} \frac{t^{\nu}}{\nu!}\bar{Q}^{\nu}(k,m)$$

når $\theta \in [\theta',\theta'']$, $a_{k-1} \leq x-(\theta-\theta') < a_k$, og $0 < t \leq \theta''-\theta$. Her defineres $\bar{P}'(x,t,\tau)$ og $\bar{Q}(k,m)$ helt analogt til $P'(x,t,n)$ og $Q(k,n)$.

Hvis vi setter

$$(4.6)\qquad P(k,m) = \sum_{\nu=0}^{\infty} \frac{(\theta''-\theta')^{\nu}}{\nu!}\bar{Q}^{\nu}(k,m),$$

blir spesielt $\bar{P}'(x,\theta''-\theta',\theta') = P(k,m)$ for alle $x \in [a_{k-1},a_k>$. Elementene $P_{ij}(k,m)$ i matrisen $P(k,m)$ er derfor de overgangssannsynlighetene man vil bruke for prognoseformål.

§ 4E. Hvis hovedhensikten med estimeringen bare er å finne estimater for $P_{ij}(k,m)$-ene, er det i grunnen ikke nødvendig å gå veien om intensitetene i det hele tatt. La nemlig $L_i(k,m)$ være antall personer i i-te område som har alder i intervallet $<a_{k-1},a_k]$ på tidspunkt θ', altså ved begynnelsen av periode m. La videre $\tilde{F}_{ij}(k,m)$ være det antall av disse som bor i område j ved utløpet av perioden, la $\tilde{U}_i(k,m)$ være det antall av de $L_i(k,m)$ personene som utvandrer i perioden, og la $\tilde{D}_i(k,m)$ være det antall av dem som dør i løpet av perioden. Da kan vi bruke estimatorene

$$(4.7)\qquad \tilde{P}_{ij}(k,m) = \tilde{F}_{ij}(k,m)/L_i(k,m) \qquad \text{for i og } j \leq N,$$

$$(4.8)\qquad \tilde{P}_{i,N+1}(k,m) = \tilde{U}_i(k,m)/L_i(k,m), \text{ og } \tilde{P}_{i,N+2}(k,m) = \tilde{D}_i(k,m)/L_i(k,m).$$

Riktignok er disse estimatorene noe mindre effisiente enn de estimatorene $\hat{P}_{ij}(k,m)$ vi kunne funnet ved hjelp av (4.4) og (4.6), siden $\tilde{P}_{ij}(k,m)$-ene ikke bruker like mye av informasjonen i dataene som $\hat{P}_{ij}(k,m)$-ene. Men effisienstapet er trolig nokså lite, og de tekniske vanskelighetene forbundet med å bruke (4.7) og (4.8) er antakelig nokså mye mindre enn dem vi støter på når vi skal beregne verdier for $\hat{P}_{ij}(k,m)$-ene.

Det er dessuten lite sannsynlig at valget mellom $\tilde{P}_{ij}(k,m)$-ene og de mer raffinerte estimatorene $\hat{P}_{ij}(k,m)$ egentlig har noen særlig betydning i forhold til de problemer vi står overfor på andre områder. For prognoseformål bør vi derfor antakelig velge den enkleste metoden og basere oss på $\tilde{P}_{ij}(k,m)$-ene.

§ 4F. Det gjenstår ennå mange problemer som må løses før resultatene ovenfor kan gis en første anvendelse i prognosemodellen. Disse problemene er knyttet til inndelingen i aldersintervaller, til oppdeling av landet i geografiske områder, samt til flyttinger innen de enkelte områder. Vi skal nevne momenter av betydning for den geografiske oppdelingen i neste kapittel, men vil ikke her gå inn på hvordan den kan foretas i praksis.

Jeg tror at vi bør ta fatt på å søke etter en slik oppdeling nokså snart.

5. Egenskaper ved områdene trekkes inn i analysen.

§ 5A. I de foregående kapitler har vi ikke eksplisitt trukket inn egenskaper ved områdene som kan ha innflytelse på flyttetilbøyeligheten, selv om effekten av dem jo vil komme frem i størrelsen av overgangsintensitetene. Vi skal nå så vidt se på hvordan analysen kan utvides til å omfatte slike egenskaper.

Vi skal ikke her prøve å fastlegge hvilke egenskaper som bør tas med i denne sammenheng, men vi kan jo antyde at vi regner med faktorer som skoletilbud, næringsgrunnlag, yrkesstruktur, inntektsgrunnlag, og kommunikasjoner som viktige. Vår fremstilling tar igjen bare sikte på å antyde det formelle grunnlag for analysen.

§ 5B. La egenskapene ved i-te område på tidspunkt θ være kvantifisert som elementer i en vektor

$$\underset{\sim}{i}(\theta) = (i_1(\theta), i_2(\theta), \ldots, i_M(\theta)).$$

La oss anta at flytteintensitetene kan uttrykkes på formen

$$\bar{\phi}_{ij}(x,\theta) = f_\theta(x, \underset{\sim}{i}(\theta), \underset{\sim}{j}(\theta)). \tag{5.1}$$

For gitt θ og x men varierende $\underset{\sim}{i}$ og $\underset{\sim}{j}$ vil $f_\theta(x, \underset{\sim}{i}, \underset{\sim}{j})$ vise hvordan flyttetilbøyeligheten varierer med egenskapene ved fra- og tilflyttingsområdet. Det er mulig at $f_\theta(x, \underset{\sim}{i}, \underset{\sim}{j})$ for gitt $x, \underset{\sim}{i}$ og $\underset{\sim}{j}$ vil variere lite med θ, dvs. at funksjonen er stabil over tiden. I så fall sløyfer vi argumentet i funksjonssymbolet, og skriver

$$\bar{\phi}_{ij}(x,\theta) = f(x, \underset{\sim}{i}(\theta), \underset{\sim}{j}(\theta)). \tag{5.2}$$

Det faktum at flytteintensiteten varierer med θ skyldes da at egenskapene ved de enkelte områder endrer seg, ikke at personene selv forandrer sin innstilling.

§ 5C. Det vil være en vanskelig men interessant oppgave å karakterisere funksjonen f. Selv om man ikke i første omgang tar sikte på å finne frem til et eksplisitt uttrykk for funksjonen, kan den bakenforliggende tankegang allikevel være til hjelp i prognosearbeidet. Hvis man tror at en teori av typen (5.2) har noe for seg, bør man jo ta dēt med i betraktningen når man velger ut de geografiske områdene for analysen i § 4E.

Nesten uansett hvordan disse områdene bygges opp, vil områdevalget i seg selv kunne tas som uttrykk for en teori om hva flyttetendensen avhenger av. En eksplisitt spesifikasjon av funksjonen f vil representere en presist formulert modell for flyttingene. Områdevalget i §§ 4E og F kan representere en mer eller mindre vagt formulert modell.

§ 5D. I § 1A nevnte vi at også egenskaper ved de N-2 andre områdene i landet har innflytelse på flyttingene $i \rightarrow j$. Det er i prinsippet mulig å trekke dette implisitt inn i analysen ved å spesifisere egenskapsvektorene for alle områdene som argumenter for funksjonen f.

6. Flytterelasjonene i 1968-versjonen av Byråets befolkningsprognosemodell.

§ 6A. Den første versjonen av den befolkningsprognosemodellen som er under arbeid i Gruppen for personmodeller, er basert på at landets ca. 450 kommuner utgjør de N områdene vi her opererer med. Modellen har følgende relasjoner:

$$(6.1) \quad L^k_{x+1}(m+1) = L^k_x(m) - D^k_x(m) + I^k_x(m) - U^k_x(m).$$

$$(6.2) \quad D^k_x(m) = q^k_x \; L^k_x(m),$$

$$(6.3) \quad U^k_x(m) = u^k_x \; L^k_x(m), \text{ og}$$

$$(6.4) \quad I^k_x(m) = i^k_x \sum_{j=1}^{N} U^j_x(m).$$

Her er $L^k_x(m)$ antall x-åringer i kommune k pr. 1. januar m, $D^k_x(m)$ er det antall av disse som dør i kommunen i år m, $U^k_x(m)$ er det antall som flytter ut av kommunen i året, og $I^k_x(m)$ er antall innflyttere i året. Man ser bort fra inn- og utvandring vis-a-vis utlandet, og likeledes ser man bort fra dødsfall blant dem som flytter i løpet av året. I tillegg til dette har modellen relasjoner for fødsler, men dem skal vi ikke trekke inn her.

§ 6B. Koeffisienten q_x^k i (6.2) representerer sannsynligheten for at en person som pr. 1. januar i et (vilkårlig) prognoseår er x år gammel og da bor i kommune k, skal dø i kommunen innen 31/12 samme år. Her gjelder

$$(6.5) \quad q_x^k = \int_0^1 \mu_k(x+t) \exp\{-\int_0^t [\mu_k(x+\xi)+\phi_k(x+\xi)]d\xi\}dt$$

når $\mu_k(x)$ representerer dødsintensiteten og $\phi_k(x)$ den totale utflytningsintensiteten[x)] i kommune k. (Vi "daterer" ikke funksjonene i dette kapitlet, for prognosemodellen opererer med funksjoner uavhengige av kalendertid.) q_x^k estimeres ved en teknikk analog til den vi beskrev i §§ 4B og C.

§ 6C. Koeffisienten u_x^k i (6.3) representerer analogt utflytningssannsynligheten. Her gjelder

$$(6.) \quad u_x^k = \int_0^1 \phi_k(x+t) \exp\{-\int_0^t [\mu_k(x+\xi)+\phi_k(x+\xi)]d\xi\}dt.$$

u_x^k estimeres som q_x^k.

Koeffisienten i_x^k i (6.4) representerer sannsynligheten for at en vilkårlig utflytter skal bo i kommune k ved utløpet av året m. Denne estimeres ved

$$(6.7) \quad \tilde{i}_x^k = \tilde{I}_x^k / \sum_{j=1}^N \tilde{U}_x^j,$$

der $\tilde{I}_x^k$ og $\tilde{U}_x^k$ representerer antall inn- og utflyttere i alder x i kommune k i et visst beregningsår (1966).

I dette regnskapet tar man ikke med eventuelle "mellomstasjoner" for personer som flytter mer enn én gang i løpet av beregningsåret. Personer som 31/12 i beregningsåret bor i samme kommune som de gjorde 1/1 samme år, regnes ikke som flyttet i det hele tatt, selv om de skulle ha flyttet ut av kommunen og så tilbake igjen i løpet av året.

§ 6D. En mulig teori bak denne prosedyren er følgende: En persons tendens til å flytte ut av den kommunen han bor i, avhenger bare av egenskaper ved denne kommunen. Antall utflyttere i et år kan da beregnes ved en relasjon som (6.3). Hvor personen bestemmer seg til å flytte hen, avhenger derimot ikke av utflyttingskommunen i det hele tatt, men bare av egenskaper ved de mulige innflyttingskommunene. Gitt at en x-årig person bestemmer seg for å flytte,

x) Sml. (3.5).

vil den betingede sannsynligheten for at han skal flytte til kommune k kunne skrives som en funksjon i_x^k som er uavhengig av hvor han bodde ved årets begynnelse. Dette gir en beregningsformel som (6.4). (Jeg vil takke Arne Øien for idéen bak disse formuleringene.)

§ 6E. I virkeligheten er folketallet i alle kommuner unntatt Oslo for lite til at man kan få skikkelige estimater for parametrene i denne modellen. Vi står også av denne grunn overfor oppgaven å danne mer passende geografiske områder.

§ 6F. Rent bortsett fra dette tror jeg at man burde innrettet seg litt annerledes ved estimeringen enn slik det er beskrevet i §§ 6B og C selv om man skulle bruke teknikken fra § 6A. Jeg ville latt $D_x^k(m)$ betegne det antall av de $L_x^k(m)$ x-åringene i kommunen 1/1 år m, som dør i løpet av året <u>uansett om de flytter før dødsfallet eller ikke</u>. Videre ville jeg latt $U_x^k(m)$ være antall utflyttere og $I_x^k(m)$ antall innflyttere <u>som overlever</u> til slutten av året. På det viset ville (6.1) til (6.4) fortsatt kunne brukes, men q_x^k, u_x^k og i_x^k ville ha et litt annet meningsinnhold enn før. Samtidig ville man slippe å "se bort fra" dødsfall blant flytterne.

Som estimatorer for de to koeffisientene i (6.2) og (6.3) ville jeg bruke (sml. § 4E.)

$$\tilde{q}_x^k = \tilde{D}_x^k/\tilde{L}_x^k \quad \text{og} \quad \tilde{u}_x^k = \tilde{U}_x^k/\tilde{L}_x^k, \tag{6.8}$$

der $\tilde{D}_x^k$, $\tilde{U}_x^k$ og $\tilde{L}_x^k$ refererer seg til en gitt beregningsperiode som det foreligger data for. Jeg ville fortsatt brukt (6.7).

7. Fordeling av innflytterne på kommunene i de enkelte områder. Flyttinger innen områdene.

§ 7A. Byråets modell skal gi prognoser for de enkelte kommuner. De områdene vi har tatt utgangspunkt i dette notatet, vil bestå av fler enn én kommune, muligens bortsett fra Oslo, som kanskje kan danne et område for seg selv. Vi blir derfor nødt til å fordele innflytterne til et område på kommunene det er sammensatt av. Likeledes må vi se på flyttingene mellom kommunene i det enkelte område. Vi skal ikke legge mye arbeid i å løse disse oppgavene her, men skal antyde hvordan man kan komme fra dem enklest mulig.

§ 7B. Innflytterne til j-te område kan fordeles proratarisk på kommunene ved at kommune k i området får en andel i_x^{jk} av de x-årige innflytterne som overlever kalenderåret. Her skal vi ha $\sum_k i_x^{jk} = 1$ for alle j og x når summen tas over kommunene i j-te område. i_x^{jk} estimeres ved en formel analog til (6,7) og (6.8).

§ 7C. Vi innfører sannsynligheten Π_x^i for at en person som bor i i-te område og har alder x ved begynnelsen av et kalenderår, skal bo i en annen kommune i samme område ved utløpet av året. Det predikerte totale antall slike"interne" x-årige flyttere i år m vil være $L_x^i(m)\Pi_x^i$, der $L_x^i(m)$ er antall x-åringer i hele området pr. 1/1 år m. Den andel av disse som fordeles til kommune k i området, kaller vi b_x^{ik}, der $\sum_k b_x^{ik} = 1$ for alle x og i. Predikert antall "interne" x-årige innflyttere til kommune k i år m blir altså $L_x^i(m)\Pi_x^i b_x^{ik}$.

Funksjonene Π_x^i og b_x^{ik} estimeres enklest mulig, dvs. analogt med (6.7) og (6.8).

§ 7D. Hvis man vil ha noen teori bak prosedyrene beskrevet i §§ 7B og C, kan man bruke et resonnement analogt til det i § 6.4.