Arbeidsnotater

STATISTISK SENTRALBYRÅ

IO 69/7 Oslo, 3. juni 1969

ANALYSE AV AVGANG FRA HETEROGENE BESTANDER MED STOKASTISKE AVGANGSINTENSITETER.

AV INGAR M. HOLME[x)]

I N N H O L D

---

x) Arbeidet er skrevet i Gruppen for personmodeller.

1. Innledning

1A. I den klassiske avgangsmodellen, slik den er beskrevet f.eks. av Sverdrup [9], forutsetter man at den bestand man betrakter, er homogen i den forstand at intensitetene for de forskjellige avgangsårsaker er felles for alle medlemmene i bestanden. I dette notatet skal vi vise hvordan den klassiske modellen kan tilpasses til en situasjon der bestanden er heterogen, slik at hvert medlem i bestanden har "sitt eget" sett av avgangsintensiteter.

La oss begynne med en kort beskrivelse av den modell vi vil modifisere: Medlemmene i en bestand på N individer kan avgå av K årsaker. Vi innfører sannsynligheten ${}_tq_x^{(v)}$ for at et x-årig medlem av bestanden skal avgå av årsak v innen alder x+t, og definerer avgangsintensitetene ved:

$$\mu_v(x) = \lim_{t\downarrow 0} {}_tq_x^{(v)}/t \qquad \text{for } v = 1,\dots,K.$$

Vi skal innskrenke oss til en situasjon der alle intensitetene er konstanter, altså der de er uavhengige av x. Bestanden observeres i ett år. Hvis individ nr. j avgår i løpet av observasjonsåret, lar vi T_j være det tilsvarende avgangstidspunktet og V_j avgangsårsaken. I motsatt fall setter vi $T_j = 1$ og $V_j = 0$. Våre observasjonsvariable er parene (T_j, V_j); $j = 1,\dots,N$. Frekvensfunksjonen for (T_j, V_j) er:

$$(1.1)\qquad f(t,v) = \begin{cases} \mu_v e^{-\mu_v} & \text{når } 0< t < 1 \text{ og } v = 1,\dots,K, \\ e^{-\mu} & \text{når } t = 1, \text{ og } v = 0, \end{cases}$$

der $\mu = \sum_{v=1}^{K} \mu_v$. La nå $D(v)$ være antall personer som avgår fra bestanden av årsak v og la $M = \sum_{i=1}^{N} T_i$ være den totale observerte levetid. Sannsynlighets-maksimeringsestimatorene for $\mu_1,\dots,\mu_K$ blir da

$$(1.2)\qquad \hat{\mu}_v = D(v)/M$$

for $v = 1,\dots,K$. Sverdrup [9] beviser at $\hat{\mu}_1,\dots,\hat{\mu}_K$ er asymptotisk uavhengige, asymptotisk normalt fordelte, asymptotisk forventningsrette og med asymptotiske varianser

$$(1.3)\qquad \text{as.var } \hat{\mu}_v = \frac{\mu_v\,\mu}{N(1-e^{-\mu})} \qquad \text{for } v = 1,\dots,K.$$

1B. Antakelsen om at intensitetene er konstante og felles for medlemmene kan ofte gjøres relativt realistisk ved at man deler opp bestanden i homogene grupper og anvender modellen på hver delbestand for seg. Med dette håper man å ta hensyn til de faktorer som er relevante for det problemet man står overfor. Slike faktorer kan være alder, sosio-økonomiske faktorer, klimatiske forhold, personenes medisinske kjennetegn, osv. For mange formål er dette godt nok. En gruppering av observasjonsmaterialet etter alle relevante kjennetegn vil imidlertid kunne resultere i at man får for få observasjoner i hver gruppe til at man kan trekke statistiske slutninger. Det kan derfor være ønskelig å finne en modell som unngår en slik sterk oppdeling ved eksplisitt å ta hensyn til heterogeniteten i bestanden.

Noen ganger kan det også hende at det overhodet ikke er mulig å dele opp bestanden fordi man ikke har observert alle de relevante faktorene hos de enkelte individer. Forfattere som Mendenhall og Hader [6], Rider [8], Swamy og Doss [10], og Cox [3] har studert modeller for slike bestander. Denne typen modeller kan beskrives på følgende måte: Anta at bestanden er sammensatt av r delbestander og la det for enkelhets skyld bare være én avgangsårsak fra hver delbestand. Avgangsintensiteten for et individ i i-te delbestand kaller vi ν_i. Hvis T er observert levetid til en person i delbestand nr. i, og hvis observasjonsperioden fortsatt er ett år, får T frekvensfunksjonen:

$$g_i(t) = \begin{cases} \nu_i e^{-\nu_i t} & \text{for } 0 < t < 1, \\ e^{-\nu_i} & \text{for } t = 1, \end{cases}$$

der i = 1,...,r. Er nå p_i sannsynligheten for at et vilkårlig valgt medlem i bestanden skal tilhøre gruppe nr. i, blir frekvensfunksjonen i levetidsfordelingen

$$g(t) = \sum_{i=1}^{r} p_i g_i(t).$$

1C. Det er lett å overbevise seg om at avgangsintensiteten for et vilkårlig valgt medlem i totalbestanden i modellen i 1B vil avhenge av t bortsett fra i trivielle spesialtilfelle, selv om avgangsintensitetene i delbestandene er konstante. I tillegg til at totalbestanden er inhomogene i den forstand at den er sammensatt av delbestander med ulike avgangsintensiteter, er altså også avgangsmodellen tids-inhomogen. Vi skal her konsentrere oss om den første typen av heterogenitet. Siden det er viktig å holde disse to formene for inhomogenitet klart fra hverandre, skal vi allikevel nevne også en annen modell som "opprinnelig" er tidshomogen, men der observasjonene kan fremstå som realisasjoner av en tids-inhomogen prosess.

Anta at en studerer en tidskontinuerlig, homogen Markov-kjede-modell og anta at en fjerner noen typer sampelstier fordi visse avgangsårsaker opptrer som "nuisance" i modellen, eller at det systematisk mangler tilgjengelig observasjonsmateriale for enkelte slags stier. Hoem [5] har bevist at den "rensede" kjeden som da fremkommer ikke nødvendigvis blir tidshomogen. Hvis statistisk analyse utføres i slike rensede kjeder, må man derfor regne med at avgangsintetensitetene kan være tidsavhengige.

1D. La oss vende tilbake til modellen i 1A, men la nå hvert enkelt medlem av bestanden ha "sitt eget" sett av avgangsintensiteter $(\mu_1,\ldots,\mu_K)$. Én måte å angripe denne situasjonen på, er å oppfatte de enkelte sett av intensiteter $(\mu_1,\ldots,\mu_K)$ som uavhengige realisasjoner av en stokastisk vektor, og tolke fordelingen gitt ved (1.1) som en betinget fordeling, gitt verdien av avgangsintensitetene. Ved å velge denne type modell, får man tatt hensyn til det forhold at variabiliteten i estimatorene $\hat{\mu}_1,\ldots,\hat{\mu}_K$ kan bestå av to hoveddeler, nemlig

(i) variasjoner som skyldes at man har et endelig sampel, og

(ii) variasjoner i intensitetene selv.

Pollard [7] nevner at variasjonsårsak (ii) kan være den viktigste når man har store bestander. Dette blir neglisjert i mange demografiske modeller i dag.

Battachayara og Holla [1] har kommet med liknende betraktninger i tilknytning til studier av levetidstesting for ulike industriprodukter. De studerer spesifikasjoner av fordelingen for $\mu_1,\ldots,\mu_K$ (for K = 1) og diskuterer blant annet hvor relevante modellene er i praktiske situasjoner. Battachayara [2] har studert ytterligere noen slike modeller. Imidlertid arbeider disse forfatterne med uavkortede levetidsfordelinger. I det etterfølgende vil vi spesielt ha i tankene personer som observasjonsobjekter. Deres avgangsintensiteter avhenger vanligvis av alderen, slik at en oppsplitting av observasjonsmaterialet i aldersgrupper er nødvendig. Vi vil derfor arbeide med avkortede fordelinger.

1E. De metoder som vi studerer i dette notatet, kan selvfølgelig anvendes i alle situasjoner der man skal foreta en avgangsanalyse for en heterogen bestand hvor man kan oppfatte én eller flere av avgangsintensitetene som stokastiske. Nærliggende demografiske eksempler er analyser av dødelighet, sykelighet, ekteskapsinngåelse, ekteskapelig fruktbarhet, og utvandring. De forfattere vi har referert til ovenfor, har beskjeftiget seg med industriell kvalitetskontroll. Man vil opplagt også kunne finne mange andre anvendelsesområder.

2. Én avgangsårsak

2A. Vi vil ta utgangspunkt i en helt enkel avgangsmodell med én avgangsårsak. N personer observeres i en periode av lengde 1 og vi lar μ være avgangsintensiteten. I tråd med hva som er sagt i 1D vil vi anta at fordelingen for observert levetid T for gitt μ-er:

$$f(t|\mu) = \begin{cases} \mu e^{-\mu t} & \text{når } 0 < t < 1, \text{ og} \\ e^{-\mu} & \text{når } t = 1, \end{cases}$$

altså en avkortet eksponensiell fordeling. For å finne den marginale fordeling for levetiden T trenger man å spesifisere en a priori-fordeling for μ. Når man skal velge en slik, er det flere hensyn å ta:

(i) Den bør ha variasjonsområde i intervallet $<0,\infty>$

(ii) Den bør være fleksibel.

(iii) Den bør være slik at den marginale levetidsfordeling blir teoretisk og numerisk "pen" å arbeide med.

Man kan tilfredsstille kravet (ii) hvis man velger en parametrisk a priori-fordeling med mange parametre. Ved å variere disse får vi frem en skare a priori-fordelinger for μ. Med et gitt observasjonsmateriale håper vi derfor at det finnes en slik fordeling som beskriver den virkelige avgang med tilstrekkelig nøyaktighet.

For å tilfredsstille hensyn (iii) bør levetidsfordelingen være "pen" i den forstand at det bør være enkelt å utføre estimering og hypoteseprøving. Kravet til numerisk enkelhet er i dag mindre strengt enn før på grunn av de muligheter moderne EDB-utstyr gir.

Hensynene (ii) og (iii) kan lett komme i konflikt med hverandre, slik at man må foreta en avveining mellom dem.

2B. Mange forslag til a priori-fordeling for μ er tenkelige. Vi har valgt å studere det tilfellet at μ er gammafordelt med tetthet

$$(2.1) \qquad g(\mu) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \mu^{\alpha-1} e^{-\beta\mu} \quad \text{for } \mu > 0;\ \alpha > 0;\ \beta > 0.$$

Den marginale fordeling for T får da frekvensfunksjonen

$$(2.2) \qquad h(t) = \int_0^{\infty} g(\mu) f(t|\mu) d\mu = \begin{cases} \alpha\ \beta^{\alpha}/(\beta+t)^{\alpha+1} & \text{når } 0 < t < 1, \\ \{\beta/(\beta+1)\}^{\alpha} & \text{når } t = 1. \end{cases}$$

Dette er en avkortet fordeling av Pareto-type. Avgangsintensiteten i fordelingen med tetthet h(t) (Hoem [4]) blir:

$$(2.3) \qquad \underset{\sim}{\mu}(t) = \frac{\alpha\ \beta^{\alpha}/(\beta+t)^{\alpha+1}}{\{\beta/(\beta+t)\}^{\alpha}} = \frac{\alpha}{\beta+t},$$

som er avhengig av t. Foruten å ta hensyn til den første type inhomogenitet nevnt i 1B, får vi altså en modell som også er tidsinhomogen. Formen på tidsavhengigheten er lite heldig. $\underset{\sim}{\mu}(t)$ avtar jo med voksende t, noe som antakelig er lite realistisk innenfor de aller fleste aldersgrupper. Oftest er imidlertid $\underset{\sim}{\mu}(t)$ av meget liten størrelsesorden slik at β må være stor i forhold til α. Da α i tillegg har tolkning som "frihetsgrader" i gammafordelingen og således formodentlig er større enn 1, kan vi regne med at $\underset{\sim}{\mu}(t)$ bare avhenger svakt av t. Dette siste "stemmer" jo forøvrig godt med intuisjonen. Da det primære i dette notatet skal være hensynet til den første type inhomogenitet (1B), vil vi si oss fornøyd selv om vi får denne lille skjevheten i modellen.

$\underset{\sim}{\mu}(t)$ er åpenbart et adekvat mål for "styrken" av avgangen. Et annet avgangsmål kunne være $E\mu = \alpha/\beta$, som er litt større enn $\underset{\sim}{\mu}(t)$ for $0 < t < 1$. Forskjellen mellom disse to målene er imidlertid formodentlig liten i de fleste tilfelle. Forøvrig ser vi at $E\mu$ har tolkning som den avgangsintensitet som gjelder på tidspunkt 0, dvs. $\underset{\sim}{\mu}(0)$.

2C. Vår oppgave er nå å estimere α og β. Vi skal først studere en situasjon der β er kjent. Det er liten grunn til å tro at β vil være kjent i praksis, men situasjonen viser seg likevel å ha en viss interesse, blant annet fordi estimeringen blir særlig enkel.).

Vi innfører $Y = \ln\{(\beta+t)/\beta\}$. Y får følgende frekvensfunksjon:

$$k(y) = \begin{cases} \alpha e^{-\alpha y} & \text{når } 0 < y < y_0, \text{ og} \\ e^{-\alpha y_0} & \text{når } y = y_0, \end{cases}$$

der $y_0 = \ln\{(\beta+1)/\beta\}$. Denne fordelingen er av samme type som i (1.1) når $k = 1$, bare med den forskjell at "observasjonsperioden" nå har en lengde på y_0. I dette tilfellet er estimeringen grei. For de N observerte personer defineres levetidene $T_1,\ldots,T_N$, $Y_j = \ln\{(\beta+T_j)/\beta\}$ og

$$D_j = \begin{cases} 1 \text{ hvis } 0 < Y_j < y_0, \\ 0 \text{ hvis } Y_j = y_0, \end{cases}$$

for $j = 1,2,\ldots,N$. Sannsynlighetsmaksimeringsestimatoren for α blir:

$$(2.4) \qquad \ddot{\alpha}^{*} = \sum_{j=1}^{N} D_j / \sum_{j=1}^{N} Y_j.$$

Som vi ser, er denne formelen helt analog med (1.2).

Når β er stor, vil

$$(2.5) \quad Y_j = \ln(1+T_j/\beta) \approx T_j/\beta.$$

Vi får da:

$$\hat{E\mu} = \alpha^*/\beta = \frac{1}{\beta}\,\frac{\Sigma D_j}{\Sigma Y_j} \approx \frac{\Sigma D_j}{\Sigma T_j}\,.$$

Det siste er jo estimatoren (1.2) med $K = 1$.

Egenskaper ved α^* kan studeres ved standardmetoder (Sverdrup [9]).

2D. La oss så se på den mer realistiske situasjon hvor også β er ukjent. Sannsynlighetsmaksimeringsestimatorene $\hat{\alpha}$ og $\hat{\beta}$ for α og β finnes nå som løsningen av følgende likningssystem:

$$(2.6) \quad \hat{\alpha} = \frac{\sum_{j=1}^{N} D_j}{\sum_{j=1}^{N} \ln(1+T_j/\hat{\beta})} \underset{\text{def.}}{\equiv} G(\hat{\beta}),$$

$$(2.7) \quad \hat{\alpha} = \frac{\hat{\beta}\sum_{j=1}^{N} D_j/(\hat{\beta}+T_j)}{N-\beta\sum_{j=1}^{N} 1/(\hat{\beta}+T_j)} \underset{\text{def.}}{\equiv} F(\hat{\beta}).$$

Dette systemet vil ha en éntydig løsning som kan finnes ved numerisk iterasjon. (Betingelsen for at iterasjonsprosessen skal konvergere er at $G'(\beta) < F'(\beta)$ nær løsningspunktet.)

La oss igjen sammenlikne estimatoren for $E\mu$ med estimatoren i (1.2) for $K = 1$. Vi bruker igjen tilnærmelsen (2.5). Likningen (2.6) gir da:

$$(2.8) \quad \hat{E\mu} = \hat{\alpha}/\hat{\beta} \approx \Sigma D_j/\Sigma T_j.$$

Da $\ln(1+T_j/\hat{\beta}) < T_j/\hat{\beta}$ såfremt $T_j > 0$, blir

$$\hat{E\mu} > \Sigma D_j/\Sigma T_j.$$

Også i tilfellet der β er ukjent vil derfor estimatoren $\hat{E\mu}$ tilnærmet være lik estimatoren i (1.2), men $\hat{E\mu}$ vil systematisk ligge litt over den "klassiske" estimator.

Sannsynlighetsmaksimeringsestimatoren for det andre avgangsmålet (2.3) blir nå:

$$\hat{\underset{\sim}{\mu}}(t) = \frac{\hat{\alpha}}{\hat{\beta}+t},$$

som vil ligge enda nærmere estimatoren i (1.2) (K = 1) for alle $t \in \langle 0,1]$ enn den i (2.8) gjør.

Asymptotiske egenskaper ved estimatorene kan finnes ut fra generell teori for sannsynlighetsmaksimeringsestimatorenes asymptotiske samplingegenskaper. Videre kan man finne tester og konfidensintervall for parametrene ved sannsynlighetskvotemetoden.

2E. Hittil har vi antatt at alle inntrådte holdes under observasjon i tidsintervallet $[0,1]$, og inntreden i bestanden i observasjonsperioden har ikke vært mulig. Man kan lempe på denne innskrenkning og anta at person nr. i har en maksimal gjenstående levetid på Z_j når han inntrer, der $Z_j \in \langle 0,1]$. Det vil da være bekvemt å oppfatte hver Z_j som en realisasjon av en stokastisk variabel (Sverdrup [9]). Man kan stille opp en modell helt analog med den vi har betraktet i dette kapitlet. Det viser seg da at formlene for estimatorene blir uforandrede, mens deres sannsynlighetsteoretiske egenskaper vil avhenge av den fordeling man spesifiserer for Z_j-ene.

3. Flere avgangsårsaker: En blandet modell

3A. Vi skal nå studere en modell med flere avgangsårsaker. Vi vil anta at μ_1 er en stokastisk variabel med sannsynlighetstetthet gitt ved (2.1), mens $\mu_2,\ldots,\mu_K$ er ukjente parametre. Forøvrig skal situasjonen være som beskrevet i 1A. (1.1) gir nå den betingede fordeling for (T_j,V_j), gitt μ_1. Den tilsvarende marginale fordeling kan skrives på formen:

$$\frac{\alpha^{d(1)}\beta^{\alpha}}{(\beta+t)^{\alpha+d(1)}} \prod_{v=2}^{K} \mu_v^{d(v)} e^{-\lambda t},$$

der $\lambda = \sum_{v=2}^{K} \mu_v$, og der $d(v) = 1$ hvis personen avgår av årsak v, $d(v) = 0$ ellers.

Vi innfører:

$$D_j(v) = \begin{cases} 1 \text{ hvis } V_j = v, \\ 0 \text{ ellers,} \end{cases}$$

for $j = 1,2,\ldots,N$ og $v = 1,2,\ldots,K$, og lar $D(v) = \sum_{j=1}^{N} D_j(v)$ for alle v. Vi vil da få sannsynlighetsmaksimeringsestimatorene i (1.2) for $\mu_2,\ldots,\mu_K$, mens sannsynlighetsmaksimeringsestimatorene for α og β finnes av (2.6) og (2.7). (Vi antar at β er ukjent.)

Det er nå lett å angi asymptotiske egenskaper ved disse estimatorene. For store N er $(\hat{\mu}_2,\ldots,\hat{\mu}_K,\hat{\alpha},\hat{\beta})$ asymptotisk multinormalt fordelt med forventning $(\mu_2,\ldots,\mu_K,\alpha,\beta)$. Den asymptotiske kovariansmatrisen har formen

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

der $A = \mathrm{diag}\,(\mu_2,\ldots,\mu_K)\cdot\frac{\lambda}{N(1-e^{-\lambda})}$ og der B er den asymptotiske kovariansmatrisen for $(\hat{\alpha},\hat{\beta})$. $\hat{\mu}_2,\ldots,\hat{\mu}_K$ er altså "som vanlig" (sml. 1A) asymptotisk uavhengige av hverandre, og de er også asymptotisk uavhengige av $(\hat{\alpha},\hat{\beta})$.

3B. Tankegangen bak denne modellen kan lett føres videre. La oss f.eks. anta at de p første av intensitetene er stokastisk uavhengige stokastiske variable, som hver for seg er gammafordelt, mens resten av intensitetene er ukjente parametre. Konklusjonene ovenfor vil da bare trenge trivielle modifikasjoner.

Forutsetningen om stokastisk uavhengighet intensitetene imellom er imidlertid ikke alltid realistisk. I mange situasjoner må man regne med at avgangsintensitetene kan være korrelerte. (Hvis en person har stor tilbøyelighet til å dø av en bestemt årsak, kan han også ha sterk tendens til å dø av andre årsaker, fordi han kanskje er så svak at han ikke makter noen særlig påkjenning i det hele tatt.) Selv om fordelingen for de p stokastiske intensitetene skulle være helt ukjent, vil allikevel sannsynlighetsmaksimeringsestimatorene for de øvrige parametriske intensitetene være gitt ved (1.2) og deres asymptotiske egenskaper vil fortsatt være som angitt i 1A.

4. To korrelerte avgangsårsaker

Som nevnt i 3B må man av og til regne med korrelasjon mellom avgangsintensitetene. Vi skal nå studere en avgangsmodell med to avgangsårsaker der begge avgangsintensitetene er stokastiske variable, som vi vil la være korrelerte.

Det er igjen mange tenkelige forslag til a priori-fordeling for (μ_1,μ_2). Vi skal ta utgangspunkt i en to-dimensjonal Wishart-fordeling, blant annet fordi marginalfordelingene da vil være gammafordelte.

Den to-dimensjonale Wishart-fordeling er sannsynlighetsfordelingen til en vektor (M_{11}, M_{12}, M_{22}) med tre elementer, og dens tetthet er:

$$f(m_{11},m_{12},m_{22}) = \frac{|M|^{(\alpha-3)/2}\exp\{-\frac{1}{2|\Sigma|}(m_{11}\beta+m_{22}\gamma-2m_{12}\rho\sqrt{\beta\gamma})\}}{2^{\alpha}\sqrt{\Pi}\ \Gamma(\frac{\alpha-1}{2})\ \Gamma(\frac{\alpha}{2})\ |\Sigma|^{\alpha/2}} ,$$

for $|M| > 0$ der $|M| = m_{11}\,m_{22} - m_{12}^2$ og $|\Sigma| = \gamma\beta(1-\rho^2)$. α, β, γ og ρ er fordelingens parametre. For å få en fordeling for to stokastiske variable, vil vi bestemme marginalfordelingen for (M_{11}, M_{22}). Vi skal så bruke denne til a-priorifordeling for (μ_1, μ_2). En del regning gir da aprioritettheten:

$$\frac{\exp\{-\frac{1}{2|\Sigma|}(\mu_1\beta+\mu_2\gamma)\}}{\sqrt{\Pi}\ 2^{\alpha}|\Sigma|^{\alpha/2}\Gamma(\frac{\alpha-1}{2})\ \Gamma(\frac{\alpha}{2})} \quad \sum_{=0}^{\infty} (\frac{\rho\sqrt{\beta\gamma}}{|\Sigma|})^{\nu}\ \frac{1}{\nu!}(\mu_1\cdot\mu_2)^{(\alpha-2+\nu)/2}\ I(\alpha,\nu),$$

der vi for $\alpha \geq 3$ har

$$I(\alpha,\nu) = \int_{-1}^{+1} (1-t^2)^{(\alpha-3)/2}t^{\nu}dt.$$

Hvis

$$s(x,y,z\ ;\alpha,\beta,\gamma,\rho) = \sum_{\nu=0}^{\infty} \frac{(\rho^{\nu}/\nu!)I(\alpha,\nu)\Gamma(\frac{\alpha+\nu}{2}+x)\Gamma(\frac{\alpha+\nu}{2}+y)}{\{\gamma(1-\rho^2)z+\frac{1}{2}\}^{(\alpha+\nu)/2+x}\{\beta(1-\rho^2)z+\frac{1}{2}\}^{(\alpha+\nu)/2+y}} ,$$

blir den tilsvarende frekvensfunksjon for (T_j, V_j)

$$\frac{\{\beta\gamma(1-\rho^2)\}^{d(1)+d(2)}s(t,d(1),d(2);\alpha,\beta,\gamma,\rho)}{2^{\alpha}\sqrt{\Pi}\ \Gamma(\frac{\alpha-1}{2})\Gamma(\frac{\alpha}{2})\beta^{\alpha/2+d(1)}\gamma^{(\alpha/2+d(2)}} .$$

Til bestemmelse av sannsynlighetsmaksimeringsestimatorene $\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\rho}$ får man likningssystemet

$$\Psi(\hat{\alpha}/2) + \Psi((\hat{\alpha}-1)/2) + \tfrac{1}{2}\ln(2\hat{\beta}\hat{\gamma}) = \frac{1}{N}\sum_{i=1}^{N}\frac{s_{i\alpha}}{s_i} ,$$

$$\hat{\alpha}/2 - \frac{1}{N}\sum_{i=1}^{N} d_i(2) = (\hat{\beta}/N)\sum_{i=1}^{N}\frac{s_{i\beta}}{s_i} ,$$

$$\hat{\alpha}/2 - \frac{1}{N}\ \Sigma d_i(1) = (\hat{\gamma}/N)\sum_{i=1}^{N}\frac{s_{i\gamma}}{s_i} ,$$

$$\frac{2\hat{\rho}}{1-\hat{\rho}^2}\sum_{i=1}^{N}[d_i(1)+d_i(2)] = \sum_{i=1}^{N}\frac{s_{i\rho}}{s_i} ,$$

der

$$s_i = s\{t_i, d_i(1), d_i(2); \hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\rho}\} \text{ og}$$

$$s_{ij} = \frac{\partial}{\partial j} s\{t_i, d_i(1), d_i(2); \hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\rho}\}$$

for $j \in \{\alpha,\beta,\gamma,\rho\}$, $i = 1,2,\ldots,N$, og der $\Psi(t) = \frac{d}{dt}\Gamma(t)$, som er en tabellert funksjon.

Jeg har ikke klart å avgjøre om dette systemet har en éntydig løsning eller om det i det hele tatt finnes noen løsning. En må derfor undersøke dette ved numeriske metoder i hvert enkelt praktisk tilfelle. Selv om det nå skulle være mulig å finne éntydige estimater for parametrene, er prosedyren komplisert og svært arbeidskrevende i praksis. Metoden synes derfor å være av begrenset praktisk verdi.

Hvis vi hadde spesifisert en annen a priorifordeling for (μ_1,μ_2), kunne resultatet blitt bedre. Trass i flere forsøk med andre fordelinger, har jeg imidlertid ikke funnet noen som ville forenkle det numeriske arbeidet i nevneverdig grad.

5. Litteraturhenvisninger:

[1] Battachayara, S. K. and M. S. Holla (1965): "On a life test distribution with stochastic deviations in the mean". Ann.Inst.Statist.Math., Tokyo 17 (1): 97-104.

[2] Battachayara, S. K. (1966): "A modified Bessel function model in life testing". Metrika 11: 133-144.

[3] Cox, D. R. (1959): "The analysis of exponentially distributed life-times with two types of failure". J.R.Statist.Soc.B 21:411-421.

[4] Hoem, Jan M. (1967): "Grunnbegreper i formell befolkningslære". Memorandum av 7/2 1967 fra Sosialøkonomisk institutt, Universitetet i Oslo.

[5] Hoem, Jan M.: "Purged and partial Markov chains". Publiseres i Skandinavisk Aktuarietidskrift.

[6] Mendenhall, W. and R. J. Hader (1958): "Estimation of parameters of mixed expnentially distributed failure time distributions from censored lifetest data". Biometrika 45: 504-520.

[7] Pollard, J. H. (1968): "A note on multi-type Galton-Watson processes with random branching probabilities". Biometrika 55 (3): 589-590.

[8] Rider, P. R. (1961): "The method of moments applied to a mixture of two exponential distributions". Ann.Math.Statist. 32:143-147.

[9] Sverdrup, Erling (1961): "Statistiske metoder ved dødelighetsundersøkelser". Stensilert. Institutt for matematiske fag, Universitetet i Oslo.

[10] Swamy, P. S. and S. A. D. C. Doss (1961): "On a problem of Bartholomew in life testing". Sankhyā A 23 (3): 225-230.