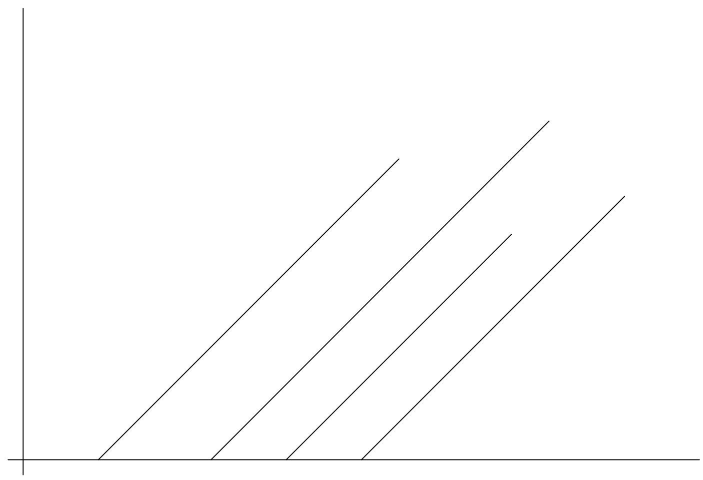
Analyse af ventetidsaksen i en fornyelsesproces og sampling i en poissonprocesmodel for Lexis diagrammet

Jens Lund

November 1996



Speciale
Københavns Universitet
Institut for Matematisk Statistik

Forsidefiguren er et Lexis diagram, der omtales i indledningen og nærmere i afsnit 2.2. Kapitel 2 og 3 omhandler generelt Lexis diagrammet.

Forord

Dette er mit speciale ved Institut for Matematisk Statistik, Københavns Universitet, for at opnå graden cand.scient. i statistik.

Martin Jacobsen, docent ved Institut for Matematisk Statistik, har været vejleder på specialet.

I appendix A er den mest brugte notation gengivet.

I appendix B og C er gengivet de nødvendige forudsætninger fra teorien om mærkede punktprocesser hhv. generelle punktprocesser. De mærkede punktprocesser bruges hovedsagelig i kapitel 1, mens kapitel 2 og 3 bruger teorien om generelle poissonprocesser.

En stor tak for hjælp og inspirerende samtaler undervejs i forløbet skal gå til min vejleder Martin Jacobsen. Bo Markussen og Helle Sørensen skal have tak for hjælp med korrekturlæsning. Niels Keiding har venligst hjulpet med at fremskaffe litteraturhenvisninger.

En kopi af dette speciale kan fås i postscript format fra
`http://www.math.ku.dk/~jlund`
og jeg kan kontaktes på e-mail adressen
`jlund@math.ku.dk`.

København, 1. nov. 1996

Jens Lund

Indhold

Indledning

Emnet for dette speciale er “flere tidsakser”. Arjas beskriver i [Arj85] tid på følgende måde:

Vi forestiller os en population af individer. Hvert individ er udstyret med en samling af stopure. Hvert stopur repræsenterer en tidsakse. Et af stopurene måler kalendertid og er derfor synkroniseret mellem individer. Resten af stopurene er typisk individuelle for hvert individ. Et stopur sættes igang, når individet fødes og måler alder af individet. Et andet stopur sættes f.eks. igang ved udbrud af sygdom og måler varigheden af sygdommen. Et tredie eksempel kunne være tid siden behandling. Ved “tid” forstå vi nu samlingen af alle stopure. Når noget sker samtidig i tid, så er det ikke på forhånd klart hvilket af stopurene der refereres til. Når vi vil analysere data fra populationen udvælger vi typisk en tidsakse, et stopur, som den centrale og lader individer have fælles nulpunkt på denne tidsakse. På denne måde flytter vi måske rundt på nogle af de andre tidsakser, der førhen var justeret efter hinanden, men ikke længere er det.

I dette speciale belyser vi to aspekter af problemet med flere tidsakser.

Kapitel 1 omhandler fornyelsesprocesser. Vi ser en fornyelsesproces over et kalendertidsinterval. En fornyelsesproces består af uafhængige identisk fordelte ventetider, som er placeret i forlængelse af hinanden. Vha. ikke-parametriske maksimum likelihood argumenter viser vi, at hvis vi ser bort fra tiden op til første fornyelse, så kan vi estimere ventetidsfordelingen som vanligt — alle ventetiderne startede til tid 0. Endvidere gennemregner vi i to konkrete eksempler egenskaberne for den tidstransformerede proces, hvor ventetiderne flyttes ned til et fælles nulpunkt. Det viser sig, at den tidstransformerede proces ikke får særligt pæne egenskaber.

Kapitel 2 og 3 omhandler Lexis diagrammet. Et Lexis diagram er et koordinatsystem, hvor vi ad den vandrette akse afbilder kalendertid og ad den lodrette akse alder. Figuren på forsiden er et Lexis diagram. Vi holder styr på de to tidsakser, kalendertid og alder, og markerer hændelser til en given kalendertid og alder som punkter i Lexis diagrammet. De to tidsakser udvikler sig med samme hastighed, og et livsforløb vil derfor blive repræsenteret af en linie med hældning 1. I kapitel 2 opstiller vi den simple model, at fødselstidspunkterne følger en poissonproces, og at levetiderne givet fødselstidspunkterne er uafhængige. Dette viser sig at føre til pæne fordelingsmæssige resultater og nemme beviser for en række sætninger. Et hovedeksempel er invalidemodellen. Vi vil hovedsagelig beskæftige os med det tidshomogene tilfælde, hvor hverken fødselsprocessen eller levetidsfordelingen afhænger af kalendertid. I kapitel 3 undersøger vi systematisk hvordan vi skal analysere levetidsfordelingen, når vi har udvalgt individer på en række angivne måder — f.eks. et tværsnit, en kohorte eller i et kalendertidsinterval. Denne ramme er nært knyttet til epidemiologi, og vi finder bl.a. den stationære rekurrenstidsfordeling og den "length biased" fordeling, som optræder i epidemiologiske sammenhænge. Den klassiske analyse med venstretrunkering og højrecensur optræder, når vi betinger med fødselstidspunkterne for de individer vi ser. Analysen er da ikke længere afhængig af antagelsen om, at fødslerne følger en poissonproces. Vi kan se det sidste som et tilfælde, hvor vi har succes med at bytte rundt på tidsakserne. Vi følger individer over kalendertid, men analyserer dem på aldersaksen.

De to dele kan læses uafhængigt af hinanden.

Resultaterne er resumeret i de tre delkonklusioner:

- Afsnit 1.9 side 28.
- Afsnit 2.8 side 45.
- Afsnit 3.13 side 74.

Appendix B og C indeholder resultater fra teorien om mærkede punktprocesser og generelle punktprocsser. De resumerer dels den teori der bruges i kapitel 1–3, dels er generelle resultater der ikke hører naturligt hjemme i et af de tre kapitler placeret i appendix. I appendix A er en oversigt over den brugte notation.

Kapitel 1

Fornyelsesprocesser

1.1 Introduktion

1.1.1 Vi vil undersøge en elpæres levetid, men har kun en fatning til rådighed. Vi starter derfor med at sætte en pære i. Når denne stopper med at virke, sætter vi straks en ny pære i og så fremdeles. Dette er et eksempel på "Testing with replacement", som f.eks. er omtalt i [ABGK93, eks. III.2.11]. Vi observerer en fornyelsesproces i kalendertid, men er interesseret i at analysere fænomenet på den tidsakse, der hedder "en pæres levetid". Vi må altså flytte alle de observerede levetider ned til et fælles nulpunkt.

Som pointeret i [Arj85] gør denne ombytning af tidsakser, at vi ikke umiddelbart kan benytte de sædvanlige martingalteknikker. Den information, der er til rådighed på de to tidsakser i et interval $[0, t]$ er forskellig; filtreringen ændres ved ombytningen af tidsakserne.

1.1.2 Hovedemnet for dette kapitel er, hvordan den ombyttede ventetidsakse opfører sig.

Specielt er vi interesseret i ikke-parametrisk estimation af den integrerede hazard $\int_0^{\cdot} \alpha$ for ventetidsfordelingen. Det vil også være dette vi mener, selv om vi siger "estimation af α".

1.1.3 I kapitlet benyttes notation og begreber fra appendix B om mærkede punktprocesser.

1.1.4 Kapitlets afsnit 1.2 og 1.3 definerer fornyelsesprocesser og hvad det vil sige, at "bytte om" på tidsakserne.

Afsnit 1.4 beskæftiger sig med estimation af α, hvor der argumenteres vha. udseendet af en likelihoodfunktion.

I afsnit 1.6 og 1.7 gennemgås to konkrete eksempler, hvor vi finder egenskaberne for den tidstransformerede proces. Hertil skal vi bruge resultater om, hvad der sker, når vi ordner n stokastiske variable efter størrelse. Disse er i afsnit 1.5.

Fornyelsesprocesser hænger nært sammen med den større klasse af semimarkov-processer. Dette vil jeg omtale i forbindelse med kommentarer til nogle artikler i litteraturen i afsnit 1.8.

Resultaterne er sammenfattet i afsnit 1.9.

1.2 Fornyelsesprocesser

1.2.1 **Definition** (Fornyelsesproces) En *fornyelsesproces* er en simpel punktproces $(T_n)_{n\geqq 1}$, hvor ventetiderne $V_n = T_n - T_{n-1}$, $n \geqq 1$, er uafhængige med samme fordeling. Sæt $T_0 \equiv 0$. En *forsinket fornyelsesproces* tillader, at V_1 har en anden fordeling end V_n, $n \geqq 2$.

1.2.2 Vi vil kun betragte kontinuerte ventetidsfordelinger (på $\mathbb{R}_+$).

1.2.3 *Bemærkning* Sammenhængen mellem den simple punktproces $(T_n)_{n\geqq 1}$, ventetiderne $(V_n)_{n\geqq 1}$ og den tilhørende tælleproces $(N_t)_{t\geqq 0}$ er

$$T_n = \sum_{i=1}^{n} V_i \qquad \text{for } n \in \mathbb{N}$$

$$N_t = \sum_{n=1}^{\infty} 1\,(T_n \leqq t) \quad \text{for } t \geqq 0$$

1.2.4 *Notation* $(\mathcal{F}_t)_{t\geqq 0}$ er den naturlige filtrering induceret af N, $\mathcal{F}_t = \sigma(N_u)_{0\leqq u\leqq t}$.

Lad i det følgende F_1 være fordelingsfunktionen for V_1, med tilhørende tæthed f_1 mht. Lebesguemålet l på $\mathbb{R}$ og hazard α_1. Tilsvarende er F, f og α fordelingen af $V_2, V_3, \ldots$ Hvis V_n, $n \in \mathbb{N}$, er identisk fordelte betegner F, f og α den fælles fordeling.

Betegn middelværdien for fordelingen F med $\mathrm{E}\,F$.

Hvis G er fordelingsfunktionen for en fordeling på $\mathbb{R}$, så er $\overline{G} = 1 - G$ den tilhørende overlevelsesfunktion.

1.2.5 Vi vil overalt i dette kapitel antage, at fordelingerne F_1 og F har endelig middelværdi.

1.2.6 **Definition** (Rekurrenstider) Hvis N er en fornyelsesproces, defineres den *bagudrettede rekurrenstid* ved $L_b(t) = t - T_{N_t}$ og den *forudrettede rekurrenstid* ved $L_f(t) = T_{N_t+1} - t$.

1.2.7 *Bemærkning* Vi vil også bruge navnene baglæns og forlæns rekurrenstider. L_b angiver tiden siden sidste spring, og L_f angiver tiden til næste spring.

1.2.8 *Bemærkning* Da tælleprocssen N_t er cadlag, bliver L_b og L_f også cadlag processer.

Filtreringen induceret af L_b er netop $(\mathcal{F}_t)$, hvorimod den forlæns rekurrenstid L_f ikke er tilpasset $(\mathcal{F}_t)$.

Ifølge [Jac95a, eks. 6.16, 6.17] bliver disse højrekontinuerte versioner af rekurrenstiderne markovprocesser mht. den filtrering de hver især inducerer. Den forlæns rekurrenstid i fordelingen givet startværdien $L_f(0) = x_0$.

1.2.9 *Notation* Tager vi grænse fra venstre til alle tider t, får vi de venstrekontinuerte versioner $L_b(t-)$ og $L_f(t-)$ af rekurrenstiderne. Per konvention er $L_b(0-) = L_b(0)$ og $L_f(0-) = L_f(0)$.

1.2.10 *Bemærkning* Da processen $(L_b(t-))_{t \geqq 0}$ er venstrekontinuert og tilpasset $(\mathcal{F}_t)$ er processen forudsigelig mht. $(\mathcal{F}_t)$. Det bliver rart, da den senere vil indgå i en intensitet.

1.2.11 *Eksempel* (Valg af V_1) Fordelingen af V_1 beskriver hvordan vi starter fornyelsesprocessen. Typisk bruges en af følgende fordelinger:

- Vi har en fornyelse til tid $t = 0$, og V_1 har samme fordeling som de øvrige ventetider.

De følgende to fordelinger er begge fra situationen, hvor vi har en *stationær* fornyelsesproces[1].

- Den stationære rekurrenstidsfordeling er fordelingen af den forudrettede rekurrenstid $L_f(t)$ til et fast tidspunkt t i en stationær fornyelsesproces. Fordelingen har tæthed $\frac{\overline{F}(x)}{\mathrm{E}\, F}$. Hvis processen er stationær får V_1 den angivne fordeling, og hvis vi lader V_1 have den pågældende fordeling bliver processen stationær. Den bagudrettede rekurrenstid $L_b(t)$ har samme stationære fordeling. Den simultane fordeling af den forlæns og baglæns rekurrenstid $(L_f(t), L_b(t)) = (X, Y)$ har tæthed $\frac{f(x+y)}{\mathrm{E}\, F}$ for $x, y \geqq 0$.

  Denne fordeling kaldes af Vardi i [Var88] for den "interception biased distribution", for at understrege, at fordelingen fremkommer som følge af måden vi sampler[2] på. F.eks. optræder fordelingen også i kapitel 3.

- Den "length biased" fordeling optræder, når vi i en stationær fornyelsesproces til et fast tidspunkt t betragter længden af det interval, der er mellem de to fornyelser, som omgiver t. Vi betragter summen $L_f(t)+L_b(t)$

[1] Kapitel XI i [Fel71] omhandler fornyelsesprocesser. Fordelingerne beskrevet her er omtalt i formel (4.6) side 369, opgave 10 side 386 og formel (4.16) side 371. "The waiting time paradox" i [Fel71, p. 12] og "An inspection paradox" i [Fel71, p. 187] er klassiske problemer, der introducerer fordelingerne her.

[2] Vi bruger det engelske ord "sample". Det danske ord "udvælge" kunne bruges, men dækker ikke på samme måde over et begreb.

af den forlæns og baglæns rekurrenstid. Fordelingen har tæthed $\frac{xdF(x)}{\mathrm{E}\,F} = \frac{xf(x)}{\mathrm{E}\,F}$. Denne fordeling ses også i kapitel 3.

De to fordelinger afhænger ikke af α på en pæn måde. Fordelingerne anvendes typisk i situationen, hvor vi har en stationær fornyelsesproces N, som vi observerer over intervallet $[t_0, t_0+t]$. Vi definerer derfor en ny fornyelsesproces $\tilde{N}$ ved $\tilde{N}(t) = N(t_0+t) - N(t_0)$, som vi observerer over $[0, t]$. Valget af F_1 fordeling for $\tilde{N}$ afhænger af hvor meget vi ser af N omkring tidspunktet t_0.

1.2.12 Lad V_1 have fordeling F_1 og V_n, $n \geqq 2$, fordeling F. Så er markovkernerne for fornyelsesprocessen N

$$
\begin{aligned}
P^{(0)}(]t,\infty]) &= e^{-\int_0^t \alpha_1(u)\,du} && \text{for } t > 0 \\
P^{(n)}_{t_1 \dots t_n}(]t,\infty]) &= e^{-\int_{t_n}^t \alpha(u-t_n)\,du} && \text{for } t > t_n
\end{aligned}
$$

Hazarden for fordelingerne $P^{(0)}$ og $P^{(n)}$ er

$$
\begin{aligned}
\alpha^{(0)}(\cdot) &= \alpha_1(\cdot) \\
\alpha^{(n)}_{t_1 \dots t_n}(\cdot) &= \alpha(\cdot - t_n)
\end{aligned}
$$

Intensitetsprocessen for fornyelsesprocessen N bliver den stokastiske proces $\lambda(t) = \alpha(L_b(t-))1\,(t > T_1) + \alpha_1(t)1\,(t \leqq T_1)$ mht. $(\mathcal{F}_t)$ og Lebesguemålet.

1.2.13 Lad τ være en stoppetid mht. en større filtrering $(\mathcal{G}_t)$, $\mathcal{F}_t \subseteq \mathcal{G}_t$. Så er højrecensureringsprocessen $C(t) = 1\,(t \leqq \tau)$, $t \geqq 0$, forudsigelig mht. $(\mathcal{G}_t)$, da den er tilpasset og venstrekontinuert. Vi vil stoppe observation af N til tid τ, dvs. højrecensurere N til tid τ.

Antag, at vi har uafhængig højrecensur, dvs. N har samme kompensator $\Lambda(\cdot) = \int_0^{\cdot} \lambda$ mht. $(\mathcal{G}_t)$ som mht. $(\mathcal{F}_t)$.

Den censurerede tælleproces N^c, $N^c(t) = N(t \wedge \tau) = \int_0^t C(u)\,dN(u)$, har intensitetsproces $\lambda^c(t) = \lambda(t)C(t)$ mht. $(\mathcal{G}_t)$ ifølge eksempel B.5.5.

1.2.14 Den observerede information op til tid t er beskrevet ved filtreringen $(\mathcal{F}^c_t)$, $\mathcal{F}^c_t = \sigma(N^c(u), C(u))_{0 \leqq u \leqq t}$.

1.2.15 *Bemærkning* $\mathcal{F}^c_t \subseteq \mathcal{G}_t$ for alle t, men vi kan ikke udtale os om $(\mathcal{F}^c_t)$ i forhold til $(\mathcal{F}_t)$.

1.2.16 *Eksempel* (Valg af τ) Følgende valg af τ er alle stoppetider mht. $(\mathcal{F}_t)$.

- $\tau = T_n$ Stop ved n'te fornyelse.
- $\tau = t_1$ Stop ved kalendertidspunkt t_1.
- $\tau = T_{N_{t_1}+1}$ Stop ved første fornyelse efter kalendertidspunkt t_1.

1.3 Ombytning af tidsakser

1.3.1 Vi observerer vores fornyelsesproces N op til stoppetiden τ, dvs. vi observerer den stoppede proces N^c og τ. Vi er interesseret i ventetidsfordelingen og vil derfor lave en tidstransformation, der flytter observationen i kalendertid over på ventetidsaksen.

1.3.2 *Notation* Tælleprocessen N og stoppetiden τ tænkes defineret på et bagvedliggende sandsynlighedesfelt $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. N er ifølge appendix B en stokastisk variabel med værdier i $(W, \mathcal{H}^\circ)$. N^c bliver også en tælleproces, og parret (N^c, τ) er en stokastisk variabel defineret på $(\Omega, \mathcal{F}, P)$ med værdier i $(W \times \mathbb{R}_0, \mathcal{H}^\circ \otimes \mathbb{B}_0)$. Filtreringen $(\mathcal{F}_t)$ er induceret af N.

1.3.3 **Definition** (af tidstransformation R til ventetidsaksen) Lad $(M, \sigma(M))$ være det mærkede punktprocesrum med mærkerum $E = \mathbb{N} \cup \{c\}$. Definer afbildningen $(N^c, \tau) \xmapsto{R} \mu$ fra $(W \times \mathbb{R}_0, \mathcal{H}^\circ \otimes \mathbb{B}_0)$ til $(M, \sigma(M))$ ved

$$\mu = \sum_{k=1}^{N_\tau^c} \varepsilon_{(T_k - T_{k-1}, k)} + 1\,(L_b(\tau) > 0)\,\varepsilon_{(L_b(\tau), c)} \tag{1.1}$$

1.3.4 *Bemærkning* Vi transformerer (N^c, τ) til en mærket punktproces, hvor tidspunkterne angiver de observerede ventetider i N^c, og mærkerne er numrene på de tilhørende ventetider, som de blev observeret på kalendertidsaksen. Det specielle mærke c er forbeholdt den ventetid, der kommer sidst i kalendertid, og som bliver censureret. Mærket c bruges kun hvis τ ikke er i et fornyelsestidspunkt.

Tidstransformationen R er illustreret i figur 1.1.

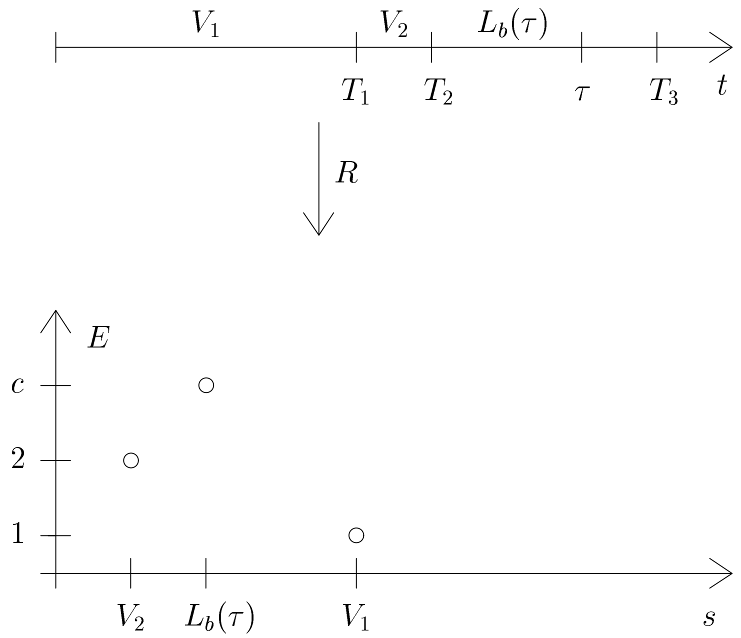
1.3.5 Antag i det følgende, at $\tau < \infty$ ns. Dette sikrer $N_\tau^c = N_\tau < \infty$ ns, og vi vil ikke have eksplosion af den tidstransformerede proces μ. Eksplosion vil sige, at der kommer uendelig mange hændelser i et endeligt tidsinterval.

1.3.6 *Bemærkning* (R målelig) Sigma-algebraen $\sigma(M)$ er frembragt af afbildningerne $\mu \mapsto \mu(A)$ for $A \in \mathbb{B}_0 \otimes \mathcal{D}(E)$. Vi skal altså betragte mængder af formen $R^{-1}(\mu([0,s] \times \{i\}) = n)$ for $(s, i, n) \in \mathbb{R}_0 \times E \times \mathbb{N}$. Disse er klart målelige og afbildningen R er målelig.

1.3.7 *Bemærkning* Indikatorfunktionen i (1.1) er nødvendig, da vi har $L_b(\tau) = 0$ med positiv sandsynlighed, hvis τ er lig en fornyelse med positiv sandsynlighed.

1.3.8 *Bemærkning* R er bijektiv som afbildning fra en mængde med sandsynlighed 1 for fordelingen af (N^c, τ) til sin billedmængde[3]. Som i bemærkning 1.3.6 kan

[3] Vi kan ikke bruge støtten som mængden med sandsynlighed 1, da vi ikke har nogen topologi på W. Vi kan derfor ikke tale om den mindste *afsluttede* mængde med sandsynlighed 1. Vi mener imidlertid blot, at vi kan rekonstruere (N^c, τ) ud fra μ, hvor μ skal ligge i "den rigtige mængde". F.eks. må intet mærke benyttes mere end én gang, og benyttes mærket $n \in \mathbb{N}$, så skal alle mærker med mindre værdi benyttes.



Figur 1.1: Illustration af tidstransformationen R. Parret (N^c, τ) transformeres over i en mærket punktproces μ med mærkerum $E = \mathbb{N} \cup \{c\}$. (Tidspunkt, mærke)-parrene for μ er angivet med en bolle.

vi indse, at den inverse afbildning R^{-1} er målelig.

1.3.9 Vi vil i de følgende afsnit interessere os for egenskaberne for den tidstransformerede proces μ. Vi vil specielt undersøge, hvilke strukturer af τ der giver en pæn struktur af μ. En pæn struktur af μ vil sige en struktur, der giver os lov til at analysere ventetiderne som vi plejer; dvs. estimere α med Nelson-Aalen estimatoren. Vi ønsker en multiplikativ struktur af det kompenserende mål for μ mht. filtreringen på $(M, \sigma(M))$ induceret af μ.

1.3.10 *Bemærkning* Selv om ventetidsfordelingen er kontinuert, er det umiddelbart klart, at det kompenserende mål for μ ikke altid har en intensitetsproces. Mærket c er problemet. Med f.eks. $\tau = T_n + 5$, og en ventetidsfordeling med støtte $\mathbb{R}_0$, vil $L_b(\tau)$ have en punktsandsynlighed i 5.

Hvis på den anden side τ er lig et fornyelsestidspunkt ns. vil $L_b(\tau) = 0$ ns. Da vi kun betragter kontinuerte ventetidsfordelinger vil jeg påstå, at μ har en intensitetsproces. Specielt er $\lambda_s^c = 0$ ns.

1.3.11 *Eksempel* $(\tau = T_n)$ Lad $n \in \mathbb{N}$ være givet og betragt $\tau = T_n$. Vi stopper observation af N ved den n'te fornyelse. Da vil ns. $N_\tau^c = n$, $L_b(\tau) = 0$ og $\mu = \sum_{k=1}^n \varepsilon_{(V_k, k)}$. Vi observerer n uafhængige ventetider, og det er velkendt, at

μ vil have intensitetsproces

$$\lambda_s^k = \begin{cases} 0 & \text{for } k \geqq n+1 \text{ og } k = c \\ \alpha_1(s) 1\,(s \leqq V_1) & \text{for } k = 1 \\ \alpha(s) 1\,(s \leqq V_k) & \text{for } 2 \leqq k \leqq n \end{cases}$$

Vi har en pæn multiplikativ struktur af intentsiteten, og vi kan bruge Nelson-Aalen estimatoren for den integrerede intensitet.

1.4 Estimation af α vha. likelihood argumenter — parametrisk og ikke-parametrisk

1.4.1 Vi betragter i dette afsnit stoppetiden $\tau = t$, og de kanoniske filtreringer på både kalendertidsaksen og den tidstransformerede akse. Den første ventetid V_1 kan have en fordeling F_1 forskellig fra resten af ventetiderne.

1.4.2 **Sætning** *Likelihoodprocessen $(L_t)_{t \geqq 0}$ på kalendertidsaksen for observation af fornyelsesprocessen N er givet ved*

$$\frac{L_t}{e^t} = \begin{cases} \overline{F}_1(t) & \textit{for } N_t = 0 \\ f_1(T_1)\overline{F}(t - T_1) & \textit{for } N_t = 1 \\ f_1(T_1) \prod_{k=2}^{N_t} \left[f(T_k - T_{k-1}) \right] \overline{F}(t - T_{N_t}) & \textit{for } N_t \geqq 2 \end{cases} \tag{1.2}$$

hvor den homogene poissonproces med parameter 1 er brugt som reference, og $T_1, \ldots, T_{N_t}$ er fornyelsestidspunkterne før tid t.

Bevis Da $P^{(n)}$ er en kontinuert fordeling på $\mathbb{R}_0$ er $P^{(n)}$ lokalt domineret af eksponentialfordelingen med parameter 1; dvs. N er lokalt domineret af den homogene poissonproces med parameter 1. (Lokal dominering: se afsnit B.6.) Ifølge 1.2.12 er intentensitetsprocessen for N

$$\lambda_t = \alpha(L_b(t-)) 1\,(t > T_1) + \alpha_1(t) 1\,(t \leqq T_1)$$

Likelihoodprocessen fås fra sætning B.6.5:

$$\begin{aligned} L_t &= \exp\left(- \left(\int_0^t (\lambda_u - 1)\, du \right) \right) \prod_{k=1}^{N_t} \frac{\lambda_{T_k}}{1} \\ &= e^t \exp\left(- \int_0^{t \wedge T_1} \alpha_1(u)\, du - \int_{T_1 \wedge t}^{t} \alpha(L_b(u))\, du \right) \prod_{k=1}^{N_t} \lambda_{T_k} \end{aligned}$$

Dette er præcis (1.2), når vi splitter ud i faktorer, der hver især vedrørerer tidsintervallet mellem to fornyelser. □

1.4.3 Vi ser, at de faktorer i likelihoodfunktionen (1.2), der svarer til tidspunkter efter T_1, er de samme som vi ville få, hvis vi observerede $N_t - 1$ levetider med hazard α og en levetid, som blev højrecensuret i tidspunktet $L_b(t)$. Her skal N_t og $L_b(t)$ opfattes som deterministiske. Hvis vi ikke siger yderligere om α, og ønsker at finde en ikke-parametrisk maksimum likelihood estimator[4] for α vha. de nævnte faktorer i likelihoodfunktionen, så viser diskussionen i [ABGK93, afsn. IV.1.5 p. 221], at vi skal benytte den sædvanlige Nelson-Aalen estimator for den integrerede intensitet $s \mapsto \int_0^s \alpha(u)\,du$. Et estimat for α selv kan f.eks. findes ved kerneudglatning. Nelson-Aalen estimatorens egenskaber som en ikke-parametrisk maksimum likelihood estimator er diskuteret i [ABGK93, afsn. VIII.4.5 p. 654].

Hvis vi ser bort fra tiden op til første fornyelse og observerer op til tid t, skal vi derfor estimere α med Nelson-Aalen estimatoren. Dette kan vi gøre, uanset om fornyelsesprocessen er stationær eller ej.

Disse argumenter giver den sædvanlige estimator for α. Imidlertid er N_t og $L_b(t)$ stokastiske, og vi kan derfor ikke benytte de sædvanlige konfidensgrænser for estimatet.

R. D. Gill bemærkede i [Gil83, p. 243] den lighed mellem likelihoodfunktionerne, som vi udnytter her. Der er yderligere kommentarer til ligheden mellem likelihoodfunktionerne i [ABGK93, eks. X.1.8].

1.4.4 *Bemærkning* Som i eksempel 1.2.11 vil fordelingen F_1 af den første ventetid ofte afhænge af α. Hvis $F_1 = F$ kan vi umiddelbart udnytte tidsrummet op til første fornyelse. Hvis $F_1 \neq F$ afhænger F_1 af α på en indviklet måde, og vi kan ikke udnytte informationen op til første fornyelse. I [Kei92, p. 318-319] er der en lignende diskussion af hvordan man skal udnytte informationen "fra starten". I [Kei92] er konklusionen, at man skal passe på antagelsen om stationaritet.

1.4.5 *Bemærkning* Hvorvidt det er afgørende at inddrage tiden op til første fornyelse i estimationen, afhænger selvfølgelig af længden af det interval, vi observerer over, i forhold til middelværdien i fordelingen F af ventetiderne.

1.4.6 Ovenstående er om ikke-parametrisk estimation af α. I [Arj85] pointeres, at hvis vi har parametriseret α ved en endelig dimensional parameter θ, så kan vi direkte benytte likelihoodprocessen (1.2) på kalendertidsaksen til at estimere θ. Endvidere kan martingalegenskaberne for likelihoodprocessen på kalendertidsaksen bruges til at finde asymptotiske resultater for $\hat{\theta}$.

Denne metode er også omtalt i [Arj86], som jeg ikke har læst, samt [ABGK93, eks. X.1.9].

1.4.7 Problemet i 1.4.3 er, at vi ikke estimerer α på ventetidsaksen, hvor α hører til. Vi har derfor ikke umiddelbart adgang til de sædvanlige asymptotiske resultater

[4] I bemærkning B.5.10 er omtalt hvad vi forstår ved en ikke-parametrisk maksimum likelihood estimator.

for Nelson-Aalen estimatoren, da den normalt motiveres ved martingalegenskaberne på den tidsakse, hvor α hører til.

Det er derfor interessant at undersøge den proces, der opstår på ventetidsaksen ved afbildningen (1.1) og se, om vi på denne tidsakse kan motivere Nelson-Aalen estimatoren for α. Dette regnes igennem for to eksempler i afsnit 1.6 og 1.7.

1.4.8 **Lemma** *Lad* $h : (E, \mathcal{E}) \to (\tilde{E}, \tilde{\mathcal{E}})$ *være en bijektiv bimålelig afbildning og betragt en domineret familie* $(P_\theta)_{\theta \in \Theta}$ *af sandsynlighedsmål på* E*;* $P_\theta = f_\theta \cdot \pi$*. Så har de transformerede sandsynlighedsmål* $h(P_\theta)$ *tæthed* $f_\theta \circ h^{-1}$ *mht.* $h(\pi)$*.*

Bevis Følger umiddelbart af [AJ91, sætn. 2 i kap. 14]. □

1.4.9 I tilfældet med $\tau = t$ er det klart, at vi ved (1.1) har en bijektiv afbildning af data på kalendertidsaksen op til tid τ og på ventetidsaksen op til tid τ. Betydningen af lemma 1.4.8 er nu, at likelihoodfunktionen er ens på de to tidsakser til tid τ. Dette lyder lovende for estimation af α. For at motivere Nelson-Aalen estimatoren som en martingalestimator på ventetidsaksen, skal vi bruge likelihoodprocessen i intervallet $[0, \tau]$ på ventetidsaksen. Da likelihoodprocesen ifølge sætning B.6.3 er en martingal mht. den kanoniske filtrering $(\mathcal{H}_s)$ på ventetidsaksen, så er $L_s = \mathrm{E}(L_\tau | \mathcal{H}_s)$ for $0 < s \leqq \tau$ på ventetidsaksen. Der er imidlertid intet, der tyder på, at det skulle være lettere at finde denne betingede middelværdi end at udføre de regninger, der kommer i afsnit 1.6 og 1.7.

1.4.10 Konklusionen på disse likelihoodbetragtninger er, at vi skal benytte den sædvanlige Nelson-Aalen estimator for α, hvis vi starter observation af vores fornyelsesproces i første fornyelse efter et fast tidspunkt.

1.5 Ordning af variable

1.5.1 Vi skal i afsnit 1.6 og 1.7 bruge resultater om ordnede variable. Vi gennemgår disse resultater her. Resultaterne udvider [AJ91, eks. 2 p. 167 og eks. 9 p. 181].

1.5.2 **Definition** (ordningsafbildning) Lad I være en endelig indexmængde, $n = |I|$, og definer *ordningsafbildningen* $h : \mathbb{R}^I \to \mathbb{R}^n$ ved $(x_i)_{i \in I} \mapsto \big(x_{(1)}, \ldots, x_{(n)}\big)$, hvor $x_{(1)} \leqq x_{(2)} \leqq \cdots \leqq x_{(n)}$ er de ordnede værdier af x_i'erne.

1.5.3 *Notation* Sættet $\big(x_{(1)}, \ldots, x_{(n)}\big)$ kaldes de ordnede observationer svarende til observationen $(x_i)_{i \in I}$. For $m = 1, \ldots, n$ kaldes $x_{(m)}$ den m'te mindste observation svarende til $(x_i)_{i \in I}$.

O_n betegner mængden $\{x \in \mathbb{R}^n | x_1 \leqq \cdots \leqq x_n\}$. O_n° er den tilsvarende åbne mængde, dvs. $O_n^\circ = \{x \in \mathbb{R}^n | x_1 < \cdots < x_n\}$. Vi har $O_1 = \mathbb{R}$.

1.5.4 *Bemærkning* $h(\mathbb{R}^n) = O_n$.

1.5.5 **Sætning** (fordeling af ordnede variable) *Lad X_i, $i \in I$, være iid med fordeling P, hvor P er en fordeling på $\mathbb{R}$ med fordelingsfunktion F og tæthed f mht. Lebesguemålet l på $\mathbb{R}$. Da gælder, at*

1. *Fordelingen af $Z = h(X) = \big(X_{(1)}, \ldots, X_{(n)}\big)$ har tæthed*

$$g(z_1, \ldots, z_n) = n! f(z_1) \cdots f(z_n) 1\left(z \in O_n^\circ\right) \tag{1.3}$$

mht. l^n.

2. *For $1 \leqq m \leqq n$ følger $X_{(m)}$ en fordeling med tæthed*

$$g(z) = n \binom{n-1}{m-1} F(z)^{m-1} f(z) \overline{F}(z)^{n-m}$$

mht. l, og fordelingsfunktion

$$G(z) = \sum_{k=m}^{n} \binom{n}{k} F(z)^k \overline{F}(z)^{n-k}.$$

Bevis Se [AJ91, eks. 9 p. 181 og eks. 2 p. 167]. □

1.5.6 *Bemærkning* Beviset for formen af tætheden (1.3) for de ordnede variable $\big(x_{(1)}, \ldots, x_{(n)}\big)$ afhænger kun af, at den simultane tæthed for $(x_i)_{i \in I}$ er symmetrisk i de n variable, og ikke af uafhængigheden. Hvis f_n er den symmetriske tæthed for $(x_i)_{i \in I}$ vil de ordnede variable $Z = \big(x_{(1)}, \ldots, x_{(n)}\big)$ have tæthed $n! f_n(z) 1\left(z \in O_n^\circ\right)$.

1.5.7 *Bemærkning* Udtrykket for hazarden for fordelingen af $X_{(m)}$ bliver ikke pænt.

1.5.8 **Sætning** (fordeling af nogle af de ordnede variable) *Lad X_i, $i \in I$, være iid med fordeling P, hvor P er en fordeling på $\mathbb{R}$ med fordelingsfunktion F og tæthed f mht. Lebesguemålet l på $\mathbb{R}$. Da gælder, at*

1. *For $1 \leqq m \leqq n$, følger $\big(X_{(1)}, \ldots, X_{(m)}\big)$ en fordeling med tæthed*

$$g(z_1, \ldots, z_m) = n^{(m)} f(z_1) \cdots f(z_m) \overline{F}(z_m)^{n-m} 1\left((z_1, \ldots, z_m) \in O_m^\circ\right)$$

mht. l^m.

2. *For $1 \leqq m \leqq n-1$, er fordelingen af $\big(X_{(m+1)}, \ldots, X_{(n)}\big)$ givet $\big(X_{(1)}, \ldots, X_{(m)}\big)$ den samme, som den ordnede fordeling af $n-m$ variable, der er uafhængige og har fordeling med tæthed $f(x)/\overline{F}(X_{(m)}) 1\big(x > X_{(m)}\big)$ mht. l.*

Bevis ved induktion efter m. For $m = 1$ har $X_{(1)}$ den angivne tæthed ifølge sætning 1.5.5. Fordelingen af $\left(X_{(2)}, \ldots, X_{(n)}\right)$ givet $X_{(1)} = z_1$ får for $z = (z_1, \ldots, z_n) \in O_n^\circ$ tæthed

$$\frac{\text{tæthed}\left(X_{(1)}, \ldots, X_{(n)}\right)}{\text{tæthed } X_{(1)}} = \frac{n! f(z_1) \cdots f(z_n)}{n f(z_1) \overline{F}(z_1)^{n-1}} = (n-1)! \frac{f(z_2)}{\overline{F}(z_1)} \cdots \frac{f(z_n)}{\overline{F}(z_1)}$$

ifølge 1.5.5, og det ønskede følger. Induktionsantagelse er, at sætningen gælder for alle k, $1 \leqq k \leqq m-1$. Lad $g(z_{k+1} | z_1, \ldots, z_k)$ betegne tætheden for $X_{(k+1)}$ givet $\left(X_{(1)}, \ldots, X_{(k)}\right) = (z_1, \ldots, z_k)$. Induktionsantagelsen, sætning 1.5.5 og succesiv betingning giver, at fordelingen af $\left(X_{(1)}, \ldots, X_{(m)}\right)$ har tæthed

$$\begin{aligned} g(z_1, \ldots, z_m) &= \prod_{k=0}^{m-1} g(z_{k+1} | z_1, \ldots, z_k) \\ &= \prod_{k=0}^{m-1} (n-k) \frac{f(z_{k+1})}{\overline{F}(z_k)} \left(\frac{\overline{F}(z_{k+1})}{\overline{F}(z_k)} \right)^{n-k-1} \\ &= n^{(m)} f(z_1) \cdots f(z_m) \overline{F}(z_m)^{n-m} \end{aligned}$$

for $z \in O_m^\circ$, hvor vi benyttede $g(z_1 | z_0) = g(z_1)$. Fordelingen af $\left(X_{(m+1)}, \ldots, X_{(n)}\right)$ givet $\left(X_{(1)}, \ldots, X_{(m)}\right)$ har nu tæthed

$$\begin{aligned} g(z_{m+1}, \ldots, z_n | z_1, \ldots, z_m) &= \frac{g(z_1, \ldots, z_n)}{g(z_1, \ldots, z_m)} \\ &= \frac{n! f(z_1) \cdots f(z_n)}{n^{(m)} f(z_1) \cdots f(z_m) \overline{F}(z_m)^{n-m}} \\ &= (n-m)! \frac{f(z_{m+1})}{\overline{F}(z_m)} \cdots \frac{f(z_n)}{\overline{F}(z_m)} \end{aligned}$$

og det ønskede følger. □

1.5.9 **Korollar** *Fordelingen af* $X_{(m+1)}$ *givet* $\left(X_{(1)}, \ldots, X_{(m)}\right) = (z_1, \ldots, z_m)$ *har tæthed*

$$(n-m) \frac{f(z)}{\overline{F}(z_m)} \left(\frac{\overline{F}(z)}{\overline{F}(z_m)} \right)^{n-m-1} 1\,(z > z_m)\,.$$

mht. l og overlevelsesfunktion

$$\overline{G}(z | z_1, \ldots, z_m) = \left(\frac{\overline{F}(z)}{\overline{F}(z_m)} \right)^{(n-m) 1 (z > z_m)}$$

1.5.10 *Bemærkning* I [Jac82, eks. 1.3.8 p. 20] bevises 1.5.9 direkte ved først at betinge yderligere med hvilke af de oprindelige variable, der er de mindste og derefter konstatere, at resultatet ikke afhænger af dette. Dvs. at den betingede sandsynlighed er målelig mht. den rigtige σ-algebra, hvor vi ikke har betinget med hvilke variable, der er de mindste.

1.6 Et tilfældigt antal ventetider

1.6.1 Vi vil i dette afsnit studere strukturen af den tidstransformerede proces (1.1), hvor vi betragter et tilfældigt antal ventetider.

1.6.2 Lad M være en stokastisk variabel med $P(M \in \mathbb{N}) = 1$ og N fornyelsesprocessen med alle ventetider iid med fordeling F. Lad yderligere M være uafhængig af N, og betragt stoppetiden $\tau = T_M$. τ er en stoppetid mht. filtreringen $(\mathcal{F}_t^N \vee \sigma(M))$ på kalendertidsaksen, og vi har uafhængig højrecensur som beskrevet i B.5.5.

1.6.3 Der vil ns. gælde, at $N_\tau = N_\tau^c = N_{T_M} = M$, $L_b(\tau) = 0$ og $\mu = \sum_{k=1}^M \varepsilon_{(V_k,k)}$. Da $L_b(\tau) = 0$, er intensiteten for censureringsmærket c lig nul, $\lambda_s^c = 0$.

1.6.4 Pga. uafhængigheden mellem N og M indeholder M ingen information om ventetidsfordelingen, og M er derfor ancilær for α. Betingningsprincippet, se f.eks. [Lau83, p. 45], siger nu, at vi skal studere egenskaberne for N i den betingede fordeling givet $M = m$. Hvis vi betinger med $M = m$, er vi i situationen fra eksempel 1.3.11, hvor vi ved, at den tidstransformerede akse giver anledning til Nelson-Aalen estimatoren vha. martingalargumenter.

På ventetidsaksen er μ og M *ikke* uafhængige, og der er derfor ikke specielt naturligt at betinge med M, når vi vil studere egenskaberne for μ. Når vi i det følgende vil forstå strukturen af tidstransformationen (1.1), vil vi betragte μ og den naturlige filtrering $(\mathcal{F}_s^\mu)$ på ventetidsaksen.

1.6.5 *Notation* $P^{(n)}$ og $\pi^{(n)}$ betegner markov- og submarkovkernerne, der beskriver fordelingen af den mærkede punktproces μ. Hazarden for $P^{(n)}$ er $\tilde{\alpha}^{(n)}$. Når det kompenserende mål har intensitetsproces, betegnes den λ_s^m, hvor $s > 0$ er tidspunktet og $m \in E$ er mærket. Tidspunkter og mærker for hændelser for den mærkede punktproces μ betegnes S_n hhv. Y_n.

Sæt $y_{\max} = \max\{y_1, \dots, y_n\}$, dvs. det hidtil største nummer på en ventetid, vi har set fejle. Afhængigheden af n i $y_{\max}$ vil fremgå af sammenhængen.

$V_{(1)}, \dots, V_{(M)}$ er de ordnede værdier af $V_1, \dots, V_M$.

1.6.6 **Sætning** ($P^{(n)}$, $\tilde{\alpha}^{(n)}$, $\pi^{(n)}$ og λ_s^m for μ) *For den tidstransformerede proces μ er for $s > s_n$*

$$P^{(n)}_{s_1 y_1 \dots s_n y_n}(]s,\infty]) = \frac{\mathrm{E}\,\overline{F}(s)^{M-n} 1\,(M \geqq n \vee y_{max})}{\mathrm{E}\,\overline{F}(s_n)^{M-n} 1\,(M \geqq n \vee y_{max})}$$

$$\tilde{\alpha}^{(n)}_{s_1 y_1 \dots s_n y_n}(s) = \alpha(s) \frac{\mathrm{E}(M-n)\overline{F}(s)^M 1\,(M \geqq n \vee y_{max})}{\mathrm{E}\,\overline{F}(s)^M 1\,(M \geqq n \vee y_{max})}$$

$$\pi^{(n)}_{s_1 y_1 \dots s_n y_n s}(y) = \begin{cases} 0 & \textit{for } y = c \\ 1\,(s \leqq V_y)\, \frac{\mathrm{E}\,\overline{F}(s)^M 1(M \geqq n+1 \vee y_{max} \vee y)}{\mathrm{E}(M-n)\overline{F}(s)^M 1(M \geqq n+1 \vee y_{max})} & \textit{for } y \in \mathbb{N} \end{cases}$$

og for $s > 0$

$$\lambda_s^y = \begin{cases} 0 & \textit{for } y = c \\ \alpha(s) 1\,(s \leqq V_y)\, \frac{\mathrm{E}\,\overline{F}(s)^M 1(M \geqq n+1 \vee y_{max} \vee y)}{\mathrm{E}\,\overline{F}(s)^M 1(M \geqq n \vee y_{max})} & \textit{for } n = \mu(]0,s[\times\mathbb{N}),\ y \in \mathbb{N} \end{cases}$$

Bevis Det antages selvfølgelig, at $s_1 y_1 \dots s_n y_n$ ligger i en passende mængde med sandsynlighed 1, dvs. samme type krav, som vi stillede i bemærkning 1.3.8. Ved beregning af $P^{(n)}$ vil vi først yderligere betinge med M. M skal ifølge [Jac95b, kap. 4 sætn. 2.9] integreres ud i fordelingen givet $S_1, Y_1, \dots, S_n, Y_n$: Med $X = (S_1 Y_1 \dots S_n Y_n)$, $Y = M$, $Z = 1\left(P^{(n)}\text{levetid i }]s,\infty]\right)$ er $\mathrm{E}(Z|X) = \mathrm{E}(\mathrm{E}(Z|X,Y)|X)$ og med $\varphi(X,Y) = \mathrm{E}(Z|X,Y)$ er iflg. [Jac95b] $\mathrm{E}(Z|X) = \mathrm{E}(\varphi(X,Y)|X) = \int \varphi(X,Y)\, P_Y(dy|X)$. Vi får

$$P^{(n)}_{s_1 y_1 \dots s_n y_n}(]s,\infty]) = \int P_{M s_1 y_1 \dots s_n y_n}(]s,\infty])\, dP(M|s_1 y_1 \dots s_n y_n) \tag{1.4}$$

Vi regner nu på de enkelte dele af (1.4). Sandsynligheden under integraltegnet skal forstås som

$$\begin{aligned} &P_{M s_1 y_1 \dots s_n y_n}(]s,\infty]) = \\ &\quad P(V_{(n+1)} > s | M = m, V_{(1)} = s_1, \dots, V_{(n)} = s_n, Y_1 = y_1, \dots, Y_n = y_n) \end{aligned} \tag{1.5}$$

Betydningen af (1.5) er, at den betingede sandsynlighed er 0, når $m < n$, da betingningen så ikke giver mening, og når $m < y_{\max}$, da $y_1 \dots y_n$ så ikke stemmer overens med hvad vi betinger med. Hvis $m = n$ giver $V_{(n+1)}$ ikke mening, men betydningen er, at der ikke skal komme flere tidspunkter; dvs. at der skal være et atom i $+\infty$, og (1.5) skal være 1. Hvis $m > n \vee y_{\max}$ er betydningen af (1.5) klar, og da der er en ligefordeling på de mulige rækkefølger ventetiderne fejler i, giver korollar 1.5.9, at udtrykket (1.5) er lig

$$\left(\frac{\overline{F}(s)}{\overline{F}(s_n)}\right)^{M-n} 1\,(M \geqq n \vee y_{\max})$$

Punktsandsynlighederne for fordelingen af M givet $s_1, y_1, \dots, s_n, y_n$ er givet ved (1.6) i det efterfølgende lemma 1.6.8. Udtrykket for $P^{(n)}$ givet ved (1.4) bliver nu

$$\begin{aligned} &P^{(n)}_{s_1 y_1 \dots s_n y_n}(]s,\infty]) \\ &= \sum_{k_1=1}^{\infty} \left(\frac{\overline{F}(s)}{\overline{F}(s_n)}\right)^{k_1-n} 1\,(k_1 \geqq n \vee y_{\max})\, \frac{p(k_1)\overline{F}(s_n)^{k_1} 1\,(k_1 \geqq n \vee y_{\max})}{\sum_{k_2=1}^{\infty} p(k_2)\overline{F}(s_n)^{k_2} 1\,(k_2 \geqq n \vee y_{\max})} \\ &= \frac{\mathrm{E}\,\overline{F}(s)^{M-n} 1\,(M \geqq n \vee y_{\max})}{\mathrm{E}\,\overline{F}(s_n)^{M-n} 1\,(M \geqq n \vee y_{\max})} \end{aligned}$$

Udtrykket for $\tilde{\alpha}^{(n)}$ fås nemt ud fra overlevelsesfunkionen ved

$$\frac{-\frac{d}{ds}P(]s,\infty])}{P(]s,\infty])}$$

$\pi^{(n)}$ findes på lignende vis. Specielt betragter vi et udtryk som (1.4) og er derfor interesseret i fordelingen af det nye mærke Y_{n+1} givet $S_1,\ldots,S_{n+1},Y_1,\ldots,Y_n$ og M. Denne er med oplagt notation

$$\pi^{(n)}_{Ms_1y_1\ldots s_ny_ns}(y)=\frac{1}{M-n}1\left(s\leqq V_y\right)1\left(1\leqq y\leqq M\right)1\left(M\geqq n+1\vee y_{\max}\right)$$

da vi har en ligefordeling på de mulige mærker. Specialt har vi $\pi^{(n)}_{\ldots}(c)=0$. Den betingede fordeling af M givet $S_1,\ldots,S_{n+1},Y_1,\ldots,Y_n$ findes præcis som i lemma 1.6.8, idet der dog er $n+1$ tidspunkter. Vi får

$$\begin{aligned}
&\pi^{(n)}_{s_1y_1\ldots s_ny_ns}(y)\\
&=\sum_{k_1=1}^{\infty}\left[\frac{1}{k_1-n}1\left(s\leqq V_y\right)1\left(k_1\geqq n+1\vee y_{\max}\vee y\right)\right.\\
&\qquad\left.\cdot\frac{p(k_1)(k_1-n)\overline{F}(s)^{k_1}1\left(k_1\geqq n+1\vee y_{\max}\right)}{\sum_{k_2=1}^{\infty}p(k_2)(k_2-n)\overline{F}(s)^{k_2}1\left(k_2\geqq n+1\vee y_{\max}\right)}\right]\\
&=1\left(s\leqq V_y\right)\frac{\mathrm{E}\,\overline{F}(s)^M1\left(M\geqq n+1\vee y_{\max}\vee y\right)}{\mathrm{E}(M-n)\overline{F}(s)^M1\left(M\geqq n+1\vee y_{\max}\right)}
\end{aligned}$$

Udtrykket for λ^y_s fås ved at gange $\tilde{\alpha}^{(n)}$ og $\pi^{(n)}$ sammen, og bemærke, at

$$(M-n)1\left(M\geqq n\vee y_{\max}\right)=(M-n)1\left(M\geqq n+1\vee y_{\max}\right)$$

□

1.6.7 *Bemærkning* Der vil ns. gælde, at $y_{\max}\geqq n$. Overalt kan vi således erstatte $n\vee y_{\max}$ med $y_{\max}$. Vi har dog oftest beholdt $n\vee y_{\max}$ for at understrege måden betingelsen fremkommer på.

Endvidere vil $y_{\max}\vee n+1=y_{\max}$, medmindre $y_{\max}=n$. Tilfældet $y_{\max}=n$ er netop når de n korteste ventetider er de n første på kalendertidsaksen.

1.6.8 **Lemma** *Fordelingen af M givet $V_{(1)}=s_1,\ldots,V_{(n)}=s_n$ og de tilhørende mærker $Y_1=y_1,\ldots,Y_n=y_n$ er givet ved punktsandsynlighederne*

$$P(M=m|s_1y_1\ldots s_ny_n)=\frac{p(m)\overline{F}(s_n)^m1\left(m\geqq n\vee y_{max}\right)}{\sum_{k=1}^{\infty}p(k)\overline{F}(s_n)^k1\left(k\geqq n\vee y_{max}\right)}\tag{1.6}$$

hvor $p(m)$ er punktsandsynlighederne i den marginale fordeling af M.

Bevis Vi vil først finde den simultane fordeling af tidspunkterne $S_1, \ldots, S_n$ og mærkerne $Y_1, \ldots, Y_n$ givet $M = m$. Dette er en fordeling på $\mathbb{R}^n \times E^n$ med tæthed mht. $l^{\otimes n} \otimes \eta^{\otimes n}$, hvor η er tællemålet på E. Bemærk først, at tætheden er 0, hvis $m < y_{\max}$ eller $m < n$. Dvs. vi skal have $m \geqq y_{\max}$ og $m \geqq n$. Endvidere er klart $(S_1, \ldots, S_n)$ og $(Y_1, \ldots, Y_n)$ uafhængige. $Y_1, \ldots, Y_n$ antager værdier i mængden af n elementer i en given rækkefølge, der er trukket uden tilbagelægning fra $\{1, \ldots, m\}$, og Y'ernes fordeling er en ligefordeling på denne mængde, som har $m^{(n)} = \frac{m!}{(m-n)!}$ elementer. Fordelingen af $S_1, \ldots, S_n$ svarer til den ordnede fordeling af de n mindste af m uafhængige variable med fordeling F, og er derfor givet ved sætning 1.5.8. Fordelingen af $S_1, \ldots, S_n, Y_1, \ldots, Y_n$ givet $M = m$ har derfor tæthed

$$m^{(n)} f(s_1) \cdots f(s_n) \overline{F}(s_n)^{m-n} \frac{1}{m^{(n)}} 1\,(m \geqq n \vee y_{\max}) = \\ f(s_1) \cdots f(s_n) \overline{F}(s_n)^{m-n} 1\,(m \geqq n \vee y_{\max})$$

for $s_1, \ldots, s_n, y_1, \ldots, y_n$ i det rigtige udfaldsrum. Den simultane fordeling med M fås ved at gange $p(m)$ på denne tæthed, der så bliver mht. $l^{\otimes n} \otimes \eta^{\otimes n} \otimes \eta_{\mathbb{N}}$ på $\mathbb{R}^n \times E^n \times \mathbb{N}$, hvor $\eta_{\mathbb{N}}$ er tællemålet på $\mathbb{N}$. Den betingede tæthed af M fås ved at normere udtrykket med summen over m; den marginale tæthed for tidspunkterne og mærkerne $S_1, \ldots, S_n, Y_1, \ldots, Y_n$. $\square$

1.6.9 *Bemærkning* Den betingede fordeling i lemma 1.6.8 afhænger af s_n. Det kan vi forstå ved følgende eksempel. Hvis F er ligefordelingen på $[0,1]$, $n = y_{\max} = 1000$, $s_n = 10^{-7}$ ville vi intuitivt forvente, at M er stor. Omvendt ville vi, hvis $s_n = 1 - 10^{-7}$, ikke forvente, at M er så meget større end n. Denne intuition stemmer med måden s_n indgår på gennem $\overline{F}(s_n)$.

1.6.10 *Bemærkning* Intensiteten λ_s^y fra sætning 1.6.6 splitter ikke op i et produkt af $\alpha(s)$ og en forudsigelig proces. Brøken med middelværdierne afhænger af α og fordelingen af M. Vi kan derfor ikke motivere Nelson-Aalen estimatoren vha. udseendet af intensitetsprocessen.

1.6.11 Det er klart, at λ_s^y er aftagende i y for $y > y_{\max}$. Når $n+1 \vee y_{\max} \vee y = n \vee y_{\max}$ reducerer λ_s^y til $\alpha(s) 1\,(s \leqq V_y)$. Dette sker netop når $y \leqq y_{\max}$ og $y_{\max} > n$, dvs. der er nogen under observation med mærke $\leqq y_{\max}$. I ord betyder dette, at fra det tidspunkt vi på den tidstransformerede akse ser mærket $\tilde{y}$, så vil alle mærker $y < \tilde{y}$ stadig under observation få den pæne intensitet $\alpha(s) 1\,(s \leqq V_y)$. For $y > y_{\max}$ er $\lambda_s^y < \alpha(s) 1\,(s \leqq V_y)$. Når vi ser et nyt $y_{\max}$ sker der således en opjustering af intensiteten for mærker $y < y_{\max}$. Dette kan forstås ved, at mærker $y < y_{\max}$ nu *skal* forekomme.

1.6.12 **Sætning** (λ_s^y som funktion af s) *Betragt $n \in \mathbb{N}$ og tilhørende y_{max}. Sæt $k = \min_{m \geqq y_{max}} \{P(M = m) > 0\}$. Da er ns*

$$\int_{s_n}^{\infty} \left(\sum_{y=k+1}^{\infty} \lambda_s^y \right) ds < \infty \tag{1.7}$$

Bevis Bemærk først, at $k < \infty$ ns. Bemærk dernæst, at $\overline{F}(s_n) > 0$ og at vi for $s > s_n$ både kan have $\overline{F}(s) = 0$ og $\overline{F}(s) > 0$. Det første tilfælde giver ikke problemer. Sæt $q = \overline{F}(s)$ og lad $p(m)$ betegne punktsandsynlighederne for den marginale fordeling af M. Vi ser nu, at

$$\begin{aligned}
\sum_{y=k+1}^{\infty} \lambda_s^y &\leqq \alpha(s) \frac{\sum_{y=k+1}^{\infty} \mathrm{E}\, q^M 1\,(M \geqq n+1 \vee y_{\max} \vee y)}{\mathrm{E}\, q^M 1\,(M \geqq n \vee y_{\max})} \\
&= \alpha(s) \frac{\sum_{y=k+1}^{\infty} \sum_{m=y}^{\infty} p(m) q^m}{\sum_{m=y_{\max}}^{\infty} p(m) q^m} \\
&= \alpha(s) q \frac{\sum_{m=k+1}^{\infty} \sum_{y=k+1}^{m} p(m) q^{m-1}}{\sum_{m=y_{\max}}^{\infty} p(m) q^m} \\
&= \alpha(s) q \frac{\sum_{m=k}^{\infty} p(m+1)(m+1-k) q^m}{\sum_{m=k}^{\infty} p(m) q^m}
\end{aligned}$$

Den første faktor $\alpha(s)\overline{F}(s)$ er en tæthed, og integralet over $]s_n, \infty]$ er derfor endeligt. Brøken kan vurderes opad med en konstant. Den er en kvotient mellem to potensrækker i q, hvor nævneren er positiv for $q > \varepsilon$. Både tæller og nævner konvergerer mod 0 for $q \to 0$. Ifølge sætning om differentiation af potensrækker, l'Hospitals regel og da $p(k) \neq 0$ vil brøken konvergere mod $\frac{k!p(k+1)(k+1-k)}{k!p(k)} = c$ for $q \to 0$. Som funktion af $s > s_n$ er brøken begrænset, da den er en kontinuert funktion på det kompakte interval $[s_n, S]$ og vil konvergere mod c for $s \to \infty$, dvs. $q \to 0$. □

1.6.13 *Bemærkning* Uligheden er en vurdering af et integral af en stokastisk proces. For de ω, hvor $k < \infty$ sker der en vurdering med en endelig grænse, som afhænger af ω. Per konvention er $\min \emptyset = \infty$ og for de ω i en nulmængde, hvor $k = \infty$, kan vi definerer, at summen i (1.7) er tom, og integralet derfor 0. Vi kan derfor sige, at vurderingen gælder overalt, og ikke blot næsten sikkert.

1.6.14 *Bemærkning* Hvis fordelingen af M har støtte på $\mathbb{N}$ er $k = y_{\max}$ ns.

1.6.15 Formel (1.7) udtrykker, at der er positiv sandsynlighed for, at vi ikke observerer mærker y med $y > k$, da fordelingen af det tilhørende springtidspunkt har et atom i $+\infty$. Dette er selvfølgelig ikke andet, end vi vil forvente fra måden μ er konstrueret på.

1.6.16 Ved at gennemgå ovenstående regninger, kan vi slække en smule på antagelsen om, at M er uafhængig af den underliggende fornyelsesproces N. Det er nok, at $P(M = m|N) = P(M = m|V_{m+1}, \dots)$. Løst sagt er M uafhængig af $V_1, \dots, V_M$. Vi har f.eks. stadig, at givet $M = m$ er $V_1, \dots, V_M$ iid med fordeling F.

1.6.17 I dette afsnit har vi betragtet en stokastisk variabel M på $\mathbb{N}$, der er uafhængig af fornyelsesprocessen N, og stoppetiden $\tau = T_M$. I denne ramme giver opførslen af den tidstransformerede proces μ ikke anledning til den sædvanlige Nelson-Aalen estimator for α, da vi ikke får en multiplikativ intensitetsmodel. Hvis vi betinger med $M = m$, er vi tilbage i eksempel 1.3.11 og betragter blot m uafhængige levetider.

1.7 Homogen poissonproces

1.7.1 Vi vil i dette afsnit betragte tilfældet $\tau = t$, hvor vi stopper observation af fornyelsesprocessen til et fast kalendertidspunkt. For at gøre det så simpelt som muligt, vil vi antage, at ventetiderne i fornyelsesprocessen alle er eksponentialfordelte med intensitet α. Fornyelsesprocessen er altså en homogen poissonproces med intensitet α.

1.7.2 Vi vil finde det kompenserende mål for μ mht. filtreringen $(\mathcal{F}_s^\mu \vee \sigma(N_t))_{s\geq 0}$, hvilket svarer til, at vi betinger med hvor mange hændelser, N_t, der er observeret i kalendertidsintervallet $[0,t]$. Det kompenserende mål for μ mht. den kanoniske filtrering $(\mathcal{F}_s^\mu)$ bliver diskuteret i 1.7.23. Bemærk, at stoppetiden τ på kalendertidsaksen er målelig mht. $\mathcal{F}_0^\mu$ på den tidstransformerede akse, og den er derfor kendt.

1.7.3 *Bemærkning* Det er velkendt, at N_t er poissonfordelt med parameter αt. Det er ligeledes velkendt, at N_t er sufficient for α ved observation af N over intervallet $[0,t]$. Det kan derfor synes underligt at betinge med N_t. Vores mål er imidlertid her i første omgang at undersøge strukturen af μ, snarere end at estimere α. Se f.eks. definition C.4.1, og [JK90, eks. 2.1 p. 173] eller korollar til sætning C.4.8 for disse påstande.

1.7.4 Vi har, at $L_b(\tau) \neq 0$ ns. Vi ser altså $N_t = n$ ventetider $V_1, \ldots, V_n$ og $L_b(\tau)$ af den sidst påbegyndte ventetid V_{n+1}, der censureres. Sæt $V_{n+1}^c = L_b(\tau)$. Vi skal interesserer os for den ordnede fordeling af $V_1, \ldots, V_n, V_{n+1}^c$, og vil kalde de ordnede værdier $V_{(1)}, \ldots, V_{(n+1)}$ uden skelnen til hvilken værdi, der blev afskåret. $V_{(1)}, \ldots, V_{(n+1)}$ er altså *ikke* de ordnede værdier af $V_1, \ldots, V_{n+1}$.

1.7.5 **Definition** (Delmængder af $\mathbb{R}^n$) Lad $y.$ betegne $\sum_{i\in I} y_i$, hvor indexmængden I er givet fra sammenhængen. Definer følgende delmængder af $\mathbb{R}^n$ og $\mathbb{R}^{n+1}$, der alle har dimension n:

$$\begin{aligned}
O_n^\circ &= \{y \in \mathbb{R}^n \,|y_1 < \cdots < y_n\} \\
O_{n,t}^\circ &= \{y \in \mathbb{R}^n \,|0 < y_1 < \cdots < y_n < t\} \\
\tilde{O}_{n,t}^\circ &= \{y \in \mathbb{R}^n \,|0 < y_1 \cdots < y_n < t - y.\} \\
H_{n+1,t} &= \{y \in \mathbb{R}^{n+1} |y_i > 0, y. = t\} \\
\tilde{H}_{n+1,t} &= \{y \in \mathbb{R}^{n+1} |0 < y_1 \cdots < y_{n+1}, y. = t\} \\
\hat{H}_{n,t} &= \{y \in \mathbb{R}^n \,|y_i > 0, y. < t\}
\end{aligned}$$

1.7.6 *Bemærkning* Vi har et linært bånd på $V_1, \dots, V_n, V_{n+1}^c$, der ligger i $H_{n+1,t}$. Dvs. at givet n af størrelserne, vil den sidste være udartet. Det vil derfor være fristende at ignorere V_{n+1}^c, og kun betragte $V_1, \dots, V_n$ som ligger i $\hat{H}_{n,t}$ og følger en fordeling på $\mathbb{R}^n$ med dimension n. I sætning 1.7.9 ser vi, at $V_1, \dots, V_n, V_{n+1}^c$ indgår fuldstændig symmetrisk, og der er derfor ikke mere grund til at udelade V_{n+1}^c fra analysen end f.eks. V_2. Vi modstår derfor fristelsen.

1.7.7 **Sætning** *Givet $N_t = n$ følger fornyelsestidspunkterne $T_1, \dots, T_n$ for N ligefordelingen på $O_{n,t}^\circ$, og hvis $U_1, \dots, U_n$ er iid ligefordelte på $[0,t]$ er $(T_1, \dots, T_n) \overset{D}{=} \big(U_{(1)}, \dots, U_{(n)}\big)$.*

Bevis Se f.eks. [JK90, p. 174] eller bemærkning C.4.5. □

1.7.8 **Korollar** *Lebesguemålet af $O_{n,t}^\circ$ er $\frac{t^n}{n!}$.*

1.7.9 **Sætning** *Givet $N_t = n$ er fordelingen af $\big(V_1, \dots, V_n, V_{n+1}^c\big)$ ligefordelingen på $H_{n+1,t}$.*

Bevis Vi finder først fordelingen af $(V_1, \dots, V_n)$, der fås ved $(V_1, \dots, V_n) = (T_1, T_2 - T_1, \dots, T_n - T_{n-1})$. Vi transformerer altså $(T_1, \dots, T_n)$, der er ligefordelt på den åbne mængde $O_{n,t}^\circ$, med en bijektiv lineær transformation med jacobideterminant 1 til den åbne mængde $\hat{H}_{n,t}$. Transformationssætningen giver nu, at $(V_1, \dots, V_n)$ er ligefordelt på $\hat{H}_{n,t}$. Afbildningen fra $\hat{H}_{n,t}$ til $H_{n+1,t}$ givet ved $(V_1, \dots, V_n) \mapsto (V_1, \dots, V_n, t - \sum_{i=1}^n V_i) = \big(V_1, \dots, V_n, V_{n+1}^c\big)$ flytter ligefordelingen på $\hat{H}_{n,t}$ over i ligefordelingen på $H_{n+1,t}$. □

1.7.10 *Bemærkning* Mængden $H_{n+1,t}$ har mål 0 mht. Lebesguemålet l^{n+1} i $\mathbb{R}^{n+1}$. Ligefordelingen på $H_{n+1,t}$ skal derfor forstås som, at vi lægger et Lebesgue mål l^n ind på det affine span af $H_{n+1,t}$ i $\mathbb{R}^{n+1}$, og betragter ligefordelingen på delmængden $H_{n+1,t}$ af denne plan.

Mere præcist kan vi tænke os en mængde $B \in \mathbb{B}_n$ med $0 < l^n(B) < \infty$ og en lineær afbildning $T : \mathbb{R}^n \to \mathbb{R}$. Vi lægger den sædvanlige ligefordeling på B med dimension n. Ligefordelingen på mængden $\{(x, Tx) | x \in B\} \subseteq \mathbb{R}^{n+1}$ defineres nu som transformationen af ligefordelingen på B ved afbildningen $x \mapsto (x, Tx)$. Tilsvarende vil vi få en sædvanlig ligefordeling af dimension n, hvis vi fra ligefordelingen på en del af en hyperplan fjerner én koordinat. Kort kan vi sige, at ligefordelingen bevares ved fordelings-dimensionsbevarende transformation. (Man kan overveje mange detaljer her. Er f.eks. disse ligefordelinger konsistente i den forstand, at man får den samme ligefordeling på hyperplanet uanset hvilke koordinater man starter med?)

1.7.11 **Sætning** *Lad $V_{(1)}, \dots, V_{(n+1)}$ være de ordnede værdier af $V_1, \dots, V_n, V_{n+1}^c$. Fordelingen af $\big(V_{(1)}, \dots, V_{(n+1)}\big)$ er ligefordelingen på $\tilde{H}_{n+1,t}$, og fordelingen af $\big(V_{(1)}, \dots, V_{(n)}\big)$ er den sædvanlige ligefordeling på $\tilde{O}_{n,t}^\circ$. Givet $\big(V_{(1)}, \dots, V_{(n)}\big)$ er $V_{(n+1)}$ udartet i $t - \sum_{i=1}^n V_{(i)}$.*

Bevis Den ordnede fordeling af $\big(V_1,\dots,V_n,V_{n+1}^c\big)$ fås ved, at planen $H_{n+1,t}$ gennemskæres af linier, der hver især bestemmer om $y_i < y_j$ eller $y_i > y_j$ for $i \neq j$. Herved inddeles planen $H_{n+1,t}$ i $(n+1)!$ felter, der hver især angiver en rækkefølge af $\big(V_1,\dots,V_n,V_{n+1}^c\big)$. Hvert af disse felter har lige stor sandsynlighed. Ordningsafbildningen flytter hver af disse felter over i $\tilde{H}_{n+1,t}$, som er et af felterne, og $\big(V_{(1)},\dots,V_{(n+1)}\big)$ følger ligefordelingen på $\tilde{H}_{n+1,t}$, da $\big(V_1,\dots,V_n,V_{n+1}^c\big)$ følger ligefordelingen på $H_{n+1,t}$.

$\big(V_{(1)},\dots,V_{(n)}\big)$ er projektionen ned på de n første koordinater af $\big(V_{(1)},\dots,V_{(n+1)}\big)$ og ligger i mængden $\tilde{O}_{n,t}^{\circ}$. Ifølge bemærkning 1.7.10 får vi igen en ligefordeling, da transformationen er fordelings-dimensionsbevarende.

Den sidste påstand i sætningen er oplagt. □

1.7.12 *Bemærkning* Da vi i bemærkning 1.7.6 insisterede på at bruge alle de variable $\big(V_1,\dots,V_n,V_{n+1}^c\big)$, har vi undervejs i de foregående argumenter mistet normeringen, der hører til ligefordelingen på $\tilde{O}_{n,t}^{\circ}$.

1.7.13 **Lemma** *Lebesguemålet af* $\tilde{O}_{n,t}^{\circ}$ *er givet ved*

$$\begin{aligned} l^n(\tilde{O}_{n,t}^{\circ}) &= \frac{1}{n}\cdot\frac{t}{n+1}\cdot l^{n-1}(\tilde{O}_{n-1,t}^{\circ}) \\ &= \frac{t^n}{n!(n+1)!} \end{aligned} \tag{1.8}$$

Bevis Følger samme teknik som i beviset for enhedskuglens Lebesgue mål i [Ber92, eks. 6.18]. Bemærk først, at $l^n(\tilde{O}_{n,t}^{\circ}) = l^n(\tilde{O}_{n,1}^{\circ})t^n$, da $\tilde{O}_{n,t}^{\circ}$ er billedet af $\tilde{O}_{n,1}^{\circ}$ ved isormorfien (bijektiv lineær afbildning) $y \mapsto ty$ med determinant t^n. For fast $y_1 \in \mathbb{R}$ er snittet af $\tilde{O}_{n,t}^{\circ}$ i $\mathbb{R}^{n-1}$ tomt, $\emptyset$, for $y_1 \notin]0,\frac{t}{n+1}[$ og er for $0 < y_1 < \frac{t}{n+1}$ bestemt ved

$$\begin{aligned} &\{(y_2,\dots,y_n)\in\mathbb{R}^{n-1} | y_1<y_2<\dots<y_n<t-\sum_{i=1}^n y_i\} = \\ &\qquad \{(y_2,\dots,y_n)\in\mathbb{R}^{n-1} | 0<y_2-y_1<\dots<y_n-y_1<t-\sum_{i=2}^n y_i-2y_1\} \end{aligned}$$

Mængden translateres ved $y_i \mapsto y_i + y_1$ (dvs. samme Lebesgue l^{n-1} mål) i

$$\{(y_2,\dots,y_n)\in\mathbb{R}^{n-1} | 0<y_2<\dots<y_n<t-\sum_{i=2}^n y_i-(n+1)y_1\}$$

Dette er præcis mængden $\tilde{O}_{n-1,t-(n+1)y_1}^{\circ}$, og udnytter vi at $l^n = l^{n-1}\otimes l$ og

Tonellis sætning, får vi

$$\begin{aligned}
l^n(\tilde{O}^\circ_{n,t}) &= \int_0^{\frac{t}{n+1}} l^{n-1}(\tilde{O}^\circ_{n-1,1})(t-(n+1)y_1)^{n-1}\,dy_1 \\
&= l^{n-1}(\tilde{O}^\circ_{n-1,1})\left[\frac{1}{n}\frac{-1}{n+1}(t-(n+1)y_1)^n\right]_0^{\frac{t}{n+1}} \\
&= l^{n-1}(\tilde{O}^\circ_{n-1,1})\frac{t^n}{n(n+1)} \\
&= \frac{1}{n}\cdot\frac{t}{n+1}\cdot l^{n-1}(\tilde{O}^\circ_{n-1,t})
\end{aligned}$$

Det lukkede udtryk for $l^n(\tilde{O}^\circ_{n,t})$ følger nemt ved induktion efter n. $\square$

1.7.14 *Bemærkning* Den rekursive version af (1.8) er inspiret af formlen for arealet af en trekant, $A = \frac{1}{2}\cdot h\cdot G$, eller i rummet formlen for rumfanget af en kegle med grundareal B, $V = \frac{1}{3}\cdot h\cdot B$.

1.7.15 *Bemærkning* Hvis vi kun havde betragtet $V_1,\dots,V_n$, og ikke $V_1,\dots,V_n,V^c_{n+1}$, ville visse argumenter være gået noget nemmere. I beviset for sætning 1.7.9 så vi, at $V_1,\dots,V_n$ var ligefordelt på $\hat{H}_{n,t}$. Specielt er tætheden $\frac{n!}{t^n}1(y\in\hat{H}_{n,t})$ symmetrisk i de n variable, og bemærkning 1.5.6 giver, at $V_{(1)},\dots,V_{(n)}$ følger en fordeling med tæthed $\frac{(n!)^2}{t^n}1\,(0<y_1<\cdots<y_n, y.<t)$. Bemærk, at Lebesguemålet af støtten for denne fordeling er en faktor $n+1$ større end $l^n(\tilde{O}^\circ_{n,t})$.

1.7.16 *Bemærkning* Den efterfølgende sætning svarer til sætning 1.5.8 med korollar.

1.7.17 **Sætning** *Lad $V_{(1)},\dots,V_{(n+1)}$ være de ordnede værdier af $V_1,\dots,V_n,V^c_{n+1}$. For $1\leqq k\leqq n$ er den betingede fordeling af $V_{(k)}$ givet $V_{(1)},\dots,V_{(k-1)}$ bestemt ved $V_{(k)}\overset{D}{=}\min_{1\leqq i\leqq n-k+1}U_i$, hvor U_i er iid ligefordelte på*

$$\left[V_{(k-1)},\frac{t-\sum_{i=1}^{k-1}V_{(i)}}{n-k+2}\right]$$

Her bruges konventionen $V_{(0)}\equiv 0$. Givet $V_{(1)},\dots,V_{(n)}$ er $V_{(n+1)}$ udartet i $t-\sum_{i=1}^n V_{(i)}$.

1.7.18 *Bemærkning* Hvis $X=\min_{1\leqq i\leqq m}U_i$, hvor U_i er iid ligefordelte på intervallet $[a,b]$, så har X overlevelsesfunktion $\overline{F}(s)=\left(\frac{b-s}{b-a}\right)^m$ for $s\in]a,b[$. Tætheden for X bliver $m\left(\frac{b-s}{b-a}\right)^{m-1}\frac{1}{b-a}1\,(x\in]a,b[)$, og hazarden er $m(b-s)^{-1}1\,(s\in]a,b[)$.

Hvis Y er betafordelt $B(1,m)$ vil $X\overset{D}{=}a+(b-a)Y$.

1.7.19 *Bevis* (for sætning 1.7.17) Påstanden om den betingede fordeling af $V_{(n+1)}$ er klar. Lad $c(y_1,\dots,c_m)$ betegne en konstant, der afhænger af $y_1,\dots,y_m$. $c(y)$ kan i det følgende indgå flere gange, og vil da ikke nødvendigvis være den samme konstant fra gang til gang. Husk at iflg. 1.7.11 er $V_{(1)}, \dots, V_{(n)}$ ligefordelt på $\tilde{O}^\circ_{n,t}$. Givet $V_{(1)} = v_1, \dots, V_{(k-1)} = v_{k-1}$ er fordelingen af $V_{(k)}, \dots, V_{(n)}$ givet ved tætheden $c(v_1, \dots, v_{k-1}) 1\big(v \in \tilde{O}^\circ_{n,t}\big)$. Bemærk, at $V_{(k)}$ i denne fordeling lever på intervallet $\big]v_{k-1}, \frac{t - \sum_{i=1}^{k-1} v_i}{n-k+2}\big[$. Øverste grænse bestemmes ved ligningen $t - \sum_{i=1}^{k-1} v_i - (n-k+1)v_k > v_k$. Vi skal nu blot integrere $V_{(k+1)}, \dots, V_{(n)}$ ud, og det interessante er indikatorfunktionen:

$$
\begin{aligned}
&\{(v_{k+1}, \dots, v_n) \in \mathbb{R}^{n-k} | 0 < v_1 < \dots < v_n < t - \sum_{i=1}^{n} v_i\} \\
&= \{(v_{k+1}, \dots, v_n) \in \mathbb{R}^{n-k} | 0 < v_{k+1} - v_k < \dots < v_n - v_k < t - \sum_{i=1}^{n} v_i - v_k\}
\end{aligned}
$$

som ved translationen $v_i \mapsto v_i + v_k$ flyttes over i

$$
\begin{aligned}
\{(v_{k+1}, \dots, v_n) \in \mathbb{R}^{n-k} | 0 < v_{k+1} < \dots < v_n < t - \sum_{i=1}^{k-1} v_i - (n-k+2)v_k - \sum_{i=k+1}^{n} v_i\} \\
= \tilde{O}^\circ_{n-k, t - \sum_{i=1}^{k-1} v_i - (n-k+2)v_k}
\end{aligned}
$$

der har Lebesguemål

$$
\frac{\Big(t - \sum_{i=1}^{k-1} v_i - (n-k+2)v_k\Big)^{n-k}}{(n-k)!(n-k+1)!}
$$

Tætheden for den betingede fordeling af $V_{(k)}$ er derfor af formen

$$
c(v_1, \dots, v_{k-1}) \Big(\frac{t - \sum_{i=1}^{k-1} v_i}{n-k+2} - s\Big)^{n-k} 1 \left(s \in \Big]v_{k-1}, \frac{t - \sum_{i=1}^{k-1} v_i}{n-k+2}\Big[\right)
$$

Da en sandsynlighedstæthed på $\mathbb{R}$ mht. l kun kan normeres på én måde, følger det ønskede nu af bemærkning 1.7.18. □

1.7.20 *Bemærkning* Sætning 1.7.17 skal kun justeres en anelse for at rumme tilfældet hvor én af de oprindelige ventetider udelades. Vi får stadig samme struktur, hvor de betingede fordelinger er minimum af nogle ligefordelinger. Det sidste tidspunkt bliver selvfølgelig ikke udartet.

1.7.21 **Korollar** *Den tidstransformerede proces μ har en intensitetsproces mht. filtreringen $(\mathcal{F}^\mu_s \vee \sigma(N_t))$ op til tidspunkt S_{N_t}. Antag $N_t = n$. Fordelingen af S_{n+1} givet $S_1, \dots, S_n$ er udartet i $t - \sum_{k=1}^{n} S_k$, og det tilhørende mærke Y_{n+1} er udartet i den ventetid, der ikke er brugt, $\{1, \dots, n, c\} \setminus \cup_{i=1}^{n} \{y_i\}$. Intensitetsprocessen* *er*

$$\lambda_s^y = \begin{cases} 0 & \textit{for } y > n \\ 1\left(s \leqq V_y\right) \frac{n-k}{n+1-k} \left(\frac{t - \sum_{i=1}^k s_i}{n-k+1} - s \right)^{-1} & \textit{for } k = \mu(]0,s[\times\{\mathbb{N} \cup c\}), \\ & y \leqq n \textit{ eller } y = c, \\ & s < S_n \end{cases}$$

Her er $1\left(s \leqq V_c\right) = 1\left(s \leqq V_{n+1}^c\right)$.

Bevis Markovkernerne $P^{(k)}_{s_1 y_1 \dots s_k y_k}$, der bestemmer fordelingen af $V_{(k+1)}$ givet de foregående tidspunkter *og* mærker, svarer til fordelingen af $V_{(k+1)}$ givet $V_{(1)}, \dots, V_{(k)}$ fra sætning 1.7.17, da fordelingen af mærkerne er uafhængig af tidspunkterne. Dette ses f.eks. fra sætning 1.7.9, hvor alle ventetiderne indgår fuldstændig symmetrisk. Dvs. mærkerne er ligefordelt på de mulige rækkefølger af $\{1, \dots, n, c\}$. Hazarden for $P^{(k)}$ fås fra sætning 1.7.17 og bemærkning 1.7.18. Herpå skal ganges $\pi^{(k)}_{s_1 y_1 \dots s_k y_k s}(y)$, hvor $\pi^{(k)}$ er ligefordelingen på de $n + 1 - k$ mærker, der ikke er brugt. □

1.7.22 Vi ser, at intensitetsprocessen for μ på den tidstransformerede akse ikke bliver pæn. Der er ikke nogen simpel pæn struktur, og resultatet synes at være meget specifikt for poissonprocessen, da det bygger kraftigt på ligefordelingen af fornyelsestidspunkterne $T_1, \dots, T_n$ fra sætning 1.7.7. Resultatet giver os derfor ikke nogen idé om, hvordan intensitetsprocessen vil se ud i tilfældet med en generel ventetidsfordeling F.

1.7.23 Det er interessant at undersøge hvad der sker, hvis vi ikke betinger med $N_t = n$. Jeg er ikke kommet frem til noget endegyldigt svar, men mener, at det bliver "grimt". En fælles komplikation i udregningerne er, at vi før eller siden skal betinge med $N_t = k$, og derfor får et problem med det stykke, som er udartet. Endvidere kommer $P^{(k)}$ f.eks. til at afhænge af de foregående tidspunkter på en ikke særlig pæn måde, se sætning 1.7.17. Man kunne f.eks. gribe sagen an på en af følgende måder:

- Med lidt upræcise regninger får man

$$\begin{aligned} & P^{(n)}_{s_1 y_1 \dots s_n y_n}(]s, \infty]) \\ & \quad = P(V_{(n+1)} > s | V_{(1)} = s_1, \dots, V_{(n)} = s_n, Y_1 = y_1, \dots, Y_n = y_n) \\ & \quad = P(\underbrace{V_{(n+1)} > s}_{*} | \underbrace{V_{(1)} = s_1, \dots, Y_n = y_n}_{**}, N_t \geqq y_{\max} \vee n - 1) \end{aligned}$$

  Den sidste linie er med for at klargøre hvad vi må forstå med de to foregående. Det er klart kun meningsgivende, hvis $N_t \geqq y_{\max}$. Endvidere må vi kræve $N_t \geqq n - 1$. Hvis $N_t = n - 1$, skal vi se n tidspunkter, og da sætter vi $V_{(n+1)} \overset{ns}{=} +\infty$, og den betingede sandsynlighed er 1. Hvis $N_t \geqq n$ tolker vi $V_{(n+1)}$ som den $n+1$'te mindste størrelse blandt $V_1, \dots, V_{N_t}, V^c_{N_t+1}$. Vi får videre, at

$$\begin{aligned} &P^{(n)}_{s_1 y_1 \dots s_n y_n}(]s,\infty]) \\ &= \frac{P(* \cap ** \cap N_t \geqq y_{\max} \vee n-1)}{P(** \cap N_t \geqq y_{\max} \vee n-1)} \\ &= \frac{\sum_k P(N_t = k) P(* \cap ** \,|N_t = k)}{\sum_k P(N_t = k) P(** \,|N_t = k)} \end{aligned}$$

og vi synes at være inde i problemet med, at vi har betinget med N_t, så der bl.a. kommer det udartede stykke.

- Vi kunne bruge metoden fra sætning 1.6.6 og lemma 1.6.8. Specielt kunne formel (1.4) med M erstattet af N_t bruges. Udtrykket for $P^{(k)}_{N_t \dots}$ under integraltegnet, hvor vi har betinget med $N_t = n$ er givet ved sætning 1.7.17 og bemærkning 1.7.18. Vi får her de 2 grimme specialtilfælde med $k = n$, hvor fordelingen er udartet i "reststykket", og $k = n+1$, hvor fordelingen er udartet i $+\infty$. Endvidere vil støtten for $P^{(k)}_{N_t \dots}$ i de pæne tilfælde afhænge af de foregående tidspunkter gennem summen. Dette harmonerer ikke særlig godt, når vi efterfølgende skal lave integrationen. Ligeledes bliver den betingede fordeling af N_t givet $S_1 Y_1 \dots S_n Y_n$ som vi skal integrere ud med hensyn til ikke pæn. Fordelingen afhænger på en kompliceret måde af de foregående tidspunkter, da vi bl.a. skal gange tætheder af formen fra bemærkning 1.7.18 sammen for at finde den simultane fordeling af $N_t, S_1, Y_1, \dots, S_n, Y_n$. Endvidere kommer igen problemet med det udartede stykke.

1.7.24 Vi har i dette afsnit undersøgt hovedeksemplet, hvor $\tau = t$. Det simpleste tilfælde, vi umiddelbart kan forestille os, er når fornyelsesprocessen er en homogen poissonproces. Dette eksempel er gennemregnet og bliver ikke pænt. Yderligere bliver resultater meget specifikt for poissonprocessen og giver ikke nogen idé til, hvordan det generelle tilfælde med ventetidsfordeling F skal se ud.

1.8 Kommentarer til litteraturen

1.8.1 Vi vil i dette afsnit kommentere nogle artikler, der beskæftiger sig med estimation i fornyelsesprocesser og de nært beslægtede semimarkovprocesser.

1.8.2 I artiklen [Var82a] beskæftiger Y. Vardi sig med situtionen, hvor vi har m observationer fra en vilkårlig fordeling F med middelværdi, og n observationer fra den tilhørende length-biased fordeling med tæthed $\frac{x dF(x)}{\mathrm{E} F}$. Han finder nu en ikke-parametrisk maksimum likelihood estimator for fordelingen F. I det tilfælde, hvor F er en absolut kontinuert fordeling, viser han passende svag konvergens af et middelværdiestimat og fordelingsfunktionen.

1.8.3 Ovenstående udnyttes bl.a. i artiklen [Var82b], hvor Vardi betragter n observationer af uafhængige stationære fornyelsesprocesser på et ikke stokastisk interval, hvor fornyelsesfordelingen F er diskret på $\mathbb{N}$. Eksistensen af en ikke-parametrisk maksimum likelihoodestimator diskuteres, og “The RT Algorithm” defineres og vises at konvergere monotont mod en ikke-parametrisk maksimum likelihoodestimator.

Undervejs er det vigtigt, at de typer af ventetider vi observerer falder i fire kategorier i forhold til sampling vinduet $[a, b]$ i fornyelsesprocessen:

1. Både start og slut ligger i vinduet. Bidrag til likelihoodfunktionen er fra fordelingen F, dvs. punktsandsynligheden $f(x)$.
2. De ufuldstændige levetider, som er først i vinduet, men slutter i vinduet. Bidrag til likelihoodfunktionen stammer fra den stationære rekurrenstidsfordeling med punktsandsynligheder $\frac{\overline{F}(x)}{\mathrm{E}\,F}$.
3. De ufuldstændige levetider, som er sidst i vinduet, men ikke først. Bidraget til likelihoodfunktionen er en højrecensureret observation fra fordelingen F, dvs. $\overline{F}(x)$.
4. De ufuldstændige levetider, der er både først og sidst i vinduet. Bidrag til likelihood observationen er en højrecensureret observation fra rekurrenstidsfordelingen, dvs. $\frac{1}{\mathrm{E}\,F}\sum_{x=b-a+1}^{\infty}\overline{F}(x)$.

Denne ramme og RT algoritmen udnytter kraftigt, at processerne er stationære, selv om RT algoritmen, der jo blot maksimaliserer en angivet (likelihood-)funktion, kan bruges i andre tilfælde; se [Var82b, afsn. 4.1]. I 1.4.3 så vi, at vi ved kun at udnytte informationen fra den først observerede fornyelse og fremad, fik en simpel Nelson-Aalen estimator for den integrerede intensitet, og at vi ikke baserede analysen på antagelsen om stationaritet. Vi udelukker altså al information, der stammer fra observationer af type 2 og 4. Vardi bruger ikke på noget tidspunkt en terminologi, der svarer til vores tidstransformation.

1.8.4 I [MD95] beskæftiger McClean og Devine sig også med ikke-parametrisk estimation i stationære fornyelsesprocesser. Til forskel fra [Var82b] er der her kun tre typer observationer. Observationer af type 4 kan vi ikke observere i den ramme, som [MD95] benytter. Det betyder, at observationer af type 2 bliver højretrunkeret i værdien $b - a =$ længden af observationsintervallet. Bidraget til likelihoodfunktionen fra denne type observation er derfor

$$\frac{\overline{F}(x)}{\mathrm{E}\,F\int_0^{b-a}\frac{\overline{F}(u)}{\mathrm{E}\,F}\,du} = \frac{\overline{F}(x)}{\int_0^{b-a}\overline{F}(u)\,du}$$

De to sidste typer af observationer giver samme bidrag til likelihoodfunktionen som hos [Var82b]. [MD95] angiver nu en estimationsmetode baseret på EM algoritmen.

1.8.5 **Definition** (Semimarkovprocesser) En markov fornyelsesproces er en proces $(X,T) = (X_i, T_j; i \in \mathbb{N}_0, j \in \mathbb{N})$, hvor X er en stokastisk proces med værdier i en tællelig mængde K og T er en proces af tidspunkter, $0 < T_1 < \cdots < T_n < \cdots < \infty$, der opfylder

$$P(X_{n+1} = k, T_{n+1} - T_n \leqq s | X_0, X_1, T_1, \dots, X_n, T_n) \\ = P(X_{n+1} = k, T_{n+1} - T_n \leqq s | X_n) \quad (1.9)$$

for $k \in K$, $s \geqq 0$.

Definer tælleprocessen N ved $N_t = \sum_{i=1}^{\infty} 1\,(T_i \leqq t)$, og den til (X,T) hørende semimarkovproces $(Y_t)_{t \geqq 0}$ er processen $Y_t = X_{N_t}$.

1.8.6 *Bemærkning* En markov fornyelsesproces bliver en fornyelsesproces, hvis der kun er en tilstand, $|K| = 1$.

1.8.7 R. D. Gill beskæftiger sig i [Gil80] med ikke-parametrisk estimation af overgangssandsynlighederne (1.9) i tilfældet, hvor tilstandsrummet K er endeligt. Først definerer Gill tælleprocesser på varighedstidsaksen i stil med vores mærkede punktproces (1.1), der tæller antal varigheder i tilstand $k \in K$ mindre end s. Derefter definerer han estimatorerne ud fra de tidstransformerede processer, som vi ville gøre, *hvis* den tidstransformerede akse gav det ideelle resultat, at tælleprocesserne her havde intensitetsproces "antal under observation"$\cdot\alpha$. Hans resultater om konsistens og svag konvergens af estimatorerne udledes herefter direkte, og beviserne bliver ret tekniske. Undervejs vises de tidstransformerede processer at have egenskaber, der minder om, hvad vi ville forvente, hvis den tidstransformerede akse gav pæne resultater. Se også [ABGK93, eks. X.1.8].

1.8.8 Hvor [Gil80] udleder sine resultater direkte, kan [VC84] udnytte martingalegenskaberne for nogle passende definerede tælleprocesser ved at indskrænke gruppen af semimarkovprocesser, de betrager.

De betragter semimarkovprocesser med endeligt mange tilstande m, hvori man kun kan bevæge sig forlæns. Dvs. at hver tilstand optræder højst en gang. Hvis tilstand i kan nås fra j, så kan j ikke nås fra i. Endvidere betragter de kun kontinuerte fordelinger. Yderligere betragter de kun det endelige tidsinterval $[0,1]$ og bemærker, at tælleprocesserne N_{ij}, der tæller antal overgange fra i til j, får intensitetsprocesser $\alpha_{ij}(t - T_i) Z_i(t)$, hvor T_i er tidspunktet for indgang i tilstand i og Z_i en passende forudsigelig proces. Deres idé er nu, at i stedet for at flytte tidsintervaller ned, så de alle starter i 0, så kan de pga. antagelserne og indskrænkningen nøjes med at strække tidsaksen $[0,1]$ til $[0,m]$, så tidsintervallet $[i-1, i[$ hører til tilstand $i \in \{1, \dots, m\}$. Endvidere kan de ved denne tidstransformation bevare martingalegenskaberne for tælleprocesserne N_{ij}.

Ved at indskrænke klassen af semimarkovprocesser opnår [VC84] væsentlig simplere beviser og teknikker end [Gil80].

1.8.9 I [ABGK93, eks. X.1.7] omtales invalidemodellen, hvor et individ er i en af de tre tilstande rask, invalid og død. Hvis man ikke kan blive rask, når man er blevet invalid, og intensiteten for overgang invalid til død kun afhænger af længden af opholdet i invalidetilstanden, så er der tale om en semimarkovproces. Man kan her foretage en tidstransformation, så indgang i invalidetilstanden regnes for tid 0 og opnå en multiplikativ intensitetsmodel for overgangen mellem invalid og død. Vi beskriver modellen i afsnit 2.5.

1.9 Konklusion

1.9.1 I dette kapitel har vi undersøgt hvordan vi skal estimere ventetidsfordelingen i en fornyelsesproces.

1.9.2 Afsnit 1.4 viste, at hvis vi opfatter Nelson-Aalen estimatoren som en ikke-parametrisk maksimum likelihood estimator, så kan vi ved at se bort fra tidsrummet fra 0 og op til første fornyelse estimere den integrerede intensitet ved Nelson-Aalen estimatoren som vi plejer. Vi skal med andre ord kun udnytte information fra første fornyelse til enden af intervallet, vi observerer N over.

Dette synes rimeligt, hvis det interval, vi observerer over, er langt i forhold til fordelingen af ventetiderne. Ydermere afhænger det ikke af nogen antagelse om stationaritet af fornyelsesprocessen.

1.9.3 Hvis ventetidsfordelingen er parametriseret, kan vi direkte udnytte likelihood-processen og martingalegenskaberne på kalendertidsaksen til at estimere og til at vise asymptotiske resultater.

1.9.4 Vi har defineret en tidstransformation (1.1), der flytter observation af en fornyelsesproces N over intervallet $[0, \tau]$ ned på den tidsakse vi kalder ventetidsaksen. I to tilfælde har vi undersøgt egenskaberne for den tidstransformerede proces i håb om, at disse var så pæne, at de motiverede Nelson-Aalen estimatoren for den integrerede intensitet. Dette var ikke tilfældet.

I det ene tilfælde, afsnit 1.6, stopper vi observation af N ved den M'te fornyelse, hvor M er uafhængig af N. Sætning 1.6.6 giver fordelingen af den tidstransformerede proces.

I det andet tilfælde, afsnit 1.7, stopper vi observation af en homogen poissonproces til tid $\tau = t$. Dette vigtige tilfælde, hvor vi stopper observation til et fast kalendertidspunkt, bliver ikke pænt, og resultatet er specifikt for den homogene poissonproces, uden umiddelbar mulighed for generalisering til en vilkårlig fornyelsesproces. Hovedresultatet er sætning 1.7.17.

1.9.5 I afsnit 1.8 omtaler vi forskellige forslag i litteraturen til estimation i fornyelsesprocesser. Det mest generelle af disse forslag, [Gil80], bliver meget teknisk i sine beviser, mens [Var82b], [MD95] og [VC84] alle udnytter en speciel struktur og estimationsmetode tilpasset deres ramme.

Kapitel 2

Poissonprocesser på Lexis diagrammet

2.1 Introduktion

2.1.1 Vi skal i dette kapitel introducere Lexis diagrammet og opstille en model for, hvordan individer bliver født og lever i Lexis diagrammet. Kapitlet drejer sig om hvilke fordelingsmæssige resultater vi kan udlede fra modellen. Det efterfølgende kapitel 3 beskæftiger sig med sampling af levetiderne i modellen.

2.1.2 Niels Keiding giver i oversigtsartiklen [Kei90] en introduktion til Lexis diagrammet og mange referencer til litteratur om forskellige former for inferens på Lexis diagrammet. Det meste af introduktionen til Lexis diagrammet stammer fra denne artikel. Lexis diagrammet er også kort omtalt i [ABGK93, eks. X.1.6].

2.1.3 I kapitlet benyttes notation og begreber fra appendix C om generelle punktprocesser — specielt poissonprocesser.

2.1.4 I de to næste, afsnit 2.2 og 2.3, definerer vi Lexis diagrammet og beskriver levetider i Lexis diagrammet med en todimensional hazard $\alpha(t, a)$.

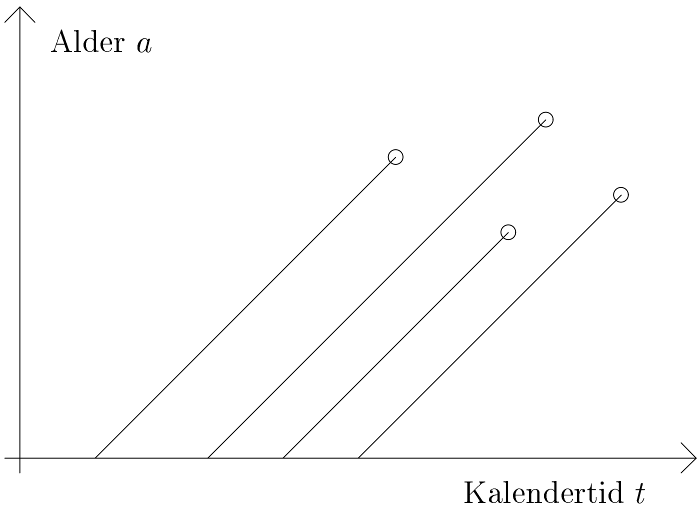
Afsnit 2.4 beskriver i detaljer modellen, hvor individer bliver født efter en poissonproces og hvor livsforløb er uafhængige givet fødselstidspunkterne. Eksemplet med en invalidemodel uden reaktivering er gennemregnet i afsnit 2.5. Disse to afsnit antyder en balancegang mellem den generelle beskrivelse af modellen og anvendelse i konkrete situationer. Dette diskuteres i afsnit 2.6.

I afsnit 2.7 omtales fire artikler, der også bruger antagelsen om, at fødslerne følger en poissonproces.

Resultaterne resumeres i afsnit 2.8.

2.2 Definition af Lexis diagrammet

2.2.1 **Definition** Et *Lexis diagram* er et koordinatsystem, hvor kalendertiden er afbildet vandret og alderen lodret. Hvert individ repræsenteres således ved en linie med hældning 1 som i figur 2.1. Dødsfald er markeret med en bolle. En bolle i punktet (t, a) betyder således, at et individ er død i alder a til kalendertid t. Dette individ er født til tid $\sigma = t - a$. Linien mellem punkterne $(\sigma, 0)$ og (t, a) vil vi kalde en *livslinie*.



Figur 2.1: Lexis diagram. Hver linie repræsenterer et livsforløb. Dødsfald er markeret med en bolle.

2.2.2 Vi taler om tre retninger i Lexis diagrammet:

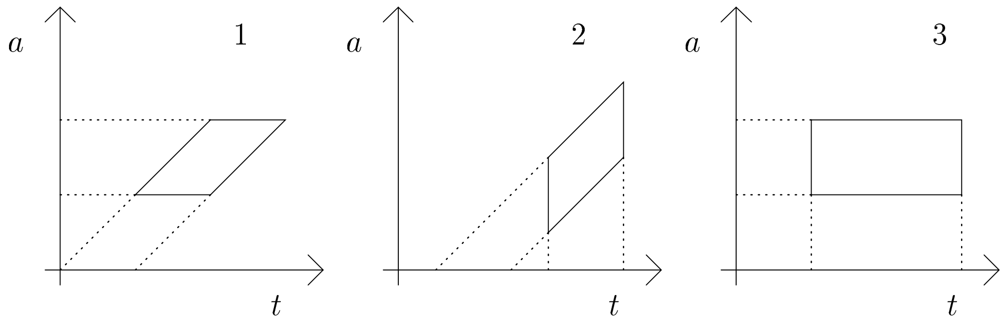
- For fast kalendertid t vil lodrette streger gå i aldersretningen.
- For fast alder a vil vandrette streger gå i kohorteretningen.
- For fast kohorte[1] t vil linier med hældning 1 gå i aldersretningen. Bemærk, at hvert år her er en faktor $\sqrt{2}$ længere end i de to andre retninger.

2.2.3 Vi taler også om tre elementærmængder for død. De tre typer mængder er illustreret i figur 2.2.

1. Et parallelogram, der angiver dødsfald mellem givne aldersgrænser for et individ født i en given periode.
2. Et parallelogram, der angiver dødsfald i et givet tidsrum for individer født i en given periode.

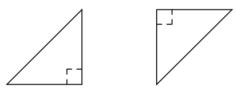
[1]En kohorte er fødselstidspunkterne i et interval.

3. Et reaktangel, der angiver dødsfald mellem givne alders- og kalendertidsgrænser.



Figur 2.2: De tre elementærmængder for død i Lexis diagrammet. 1 angiver dødsfald mellem aldersgrænser for individer født i en given periode. 2 angiver dødsfald i en given kalendertidsperiode for individer født i en given periode. 3 angiver dødsfald i en given kalendertidsperiode mellem givne aldersgrænser.

2.2.4 Elementærmængderne for død er opbygget af ligebenede trekanter hvor topvinklen er ret, og med de to mulige orienteringer, som er skitseret i figur 2.3.



Figur 2.3: De to typer elementærtrekanter i Lexis diagrammet, som elementærmængderne for død er opbygget af.

2.2.5 *Bemærkning* Så længe forholdet mellem længden af de to tidsintervaller, der definerer en elementærmængde for død i Lexisdiagrammet, er et naturligt tal, så er det umiddelbart klart, at elementærmængden kan sammenstykkes af elementærtrekanter — endda af samme størrelse. Det er denne situation, vi vil støde på i praksis.

Generelt vil jeg imidlertid mene, at det på ingen måde er klart, at elementærmængderne for død kan sammenstykkes af endeligt mange elementærtrekanter. Jeg vil påstå, at

> En elementærmængde for død i Lexis diagrammet kan sammenstykkes af endeligt mange elementærtrekanter, hvis og kun hvis forholdet mellem længden af de to tidsintervaller, der definerer elementærmængden, er rationalt.

Man kunne nu overveje det mindste antal elementærtrekanter, der skal bruges og i hvor mange størrelser ... Det er af symmetrigrunde oplagt, at når man bruger færrest mulige elementærtrekanter til at sammenstykke en given elementærmængde, så vil man bruge lige mange elementærtrekanter af hver type.

2.2.6 *Bemærkning* I en situation, hvor vi beskriver hvert individs livsforløb med en mærket punktproces, kan vi afbilde det ved at afsætte mærkerne på livslinien i figur 2.1.

2.3 Beskrivelse af levetider på Lexis diagrammet

2.3.1 Vi vil antage, at individers levetider er bestemt af dødsintensiteten $\alpha(t,a)$, hvor første argument er kalendertiden t og andet argument er alderen a. Fortolkningen af α er, at for h lille og et individ født til tid s er $\alpha(s+a,a)h$ sandsynligheden for at individet dør i aldersintervallet $]a,a+h]$, givet at individet mindst bliver alder a. Hvis X er levetiden for et individ født til tid s er altså $P(a < X \leqq a+h | X \geqq a) \approx \alpha(s+a,a)h$. Mere præcist er overlevelsesfunktionen $\overline{F}_s$ for et individ med levetid X født til kalendertid s

$$\overline{F}_s(x) = P_s(X > x) = \exp\left(-\int_0^x \alpha(s+a,a)\,da\right)$$

Kald den tilhørende tæthed f_s, $f_s(a) = \overline{F}_s(a)\alpha(s+a,a)$.

2.3.2 *Eksempel* (Stykkevis konstante intensiteter) Vi får en parametrisk model for α, hvis vi lader α være stykkevis konstant. Når α er stykkevis konstant, inddeles (t,a)-planen i disjunkte områder, f.eks. kvadrater, hvis forening giver hele planen. α er konstant på hvert af disse områder. Præcis som i det sædvanlige tilfælde med én tidsakse bliver estimatet for hver af parametrene en "occurence exposure rate" — antal dødsfald i området divideret med det totale tidsrum levet i området.

Det er muligt at inddrage baggrundsvariable i analysen. Hvis deres virkning på intensiteten er log-lineær, kan man bruge programmer fra log-linære poissonmodeller, til at estimere parametrene og foretage test.

Denne analyse holder under antagelsen, at levetiderne for individer er uafhængige givet fødselstidspunkterne.

Keiding diskuterer i [Kei90] problemet med at identificere parametrene i en opspalting af intensiteten i alders, periode og kohorte effekter. Der er også yderligere referencer til dette problem.

2.4 Modellen — fødsler følger en poissonproces og levetider er uafhængige

2.4.1 Antag, at $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$ er en poissonproces på $\mathbb{R}$ med intensitet $\beta(t)$, $t \in \mathbb{R}$. Kald intensitetsmålet for Φ, $\Phi = \beta \cdot l$. η er en poissonproces i den forstand, som definition C.4.1 giver. Intensiteten β skal altså opfylde, at for alle endelige intervaller $[t_1, t_2]$ er $\int_{t_1}^{t_2} \beta(s)\, ds < \infty$. Poissonprocessen η beskriver fødselstidspunkterne på kalendertidsaksen og vil i det følgende blive kaldt *fødselsprocessen*.

2.4.2 Givet fødselsprocessen $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$ vil vi til hvert fødselstidspunkt σ_i knytte en levetid X_i. Givet fødselstidspunkterne skal levetiderne være uafhængige og deres fordeling kun afhænge af fødselstidspunktet σ_i for levetiden X_i, samt eventuelle parametre.

2.4.3 Lad P være en markovkerne fra et måleligt rum $(S, \mathbb{S})$ til et andet måleligt rum $(E, \mathcal{E})$. P er en afbildning $P : S \times \mathcal{E} \to [0, 1]$, der opfylder

- For alle $s \in S$ er $P(s, \cdot)$ et sandsynlighedsmål på $(E, \mathcal{E})$.
- For alle $A \in \mathcal{E}$ er afbildningen $s \mapsto P(s, A)$ fra S til $\mathbb{R}$ målelig.

Vi vil undertiden skrive $P_s(A)$ for $P(s, A)$.

2.4.4 **Sætning** *Lad η være en poissonproces på $(S, \mathbb{S})$ med intensitetsmål Φ. Lad der yderligere være givet et rum $(E, \mathcal{E})$ og en markovkerne P fra $(S, \mathbb{S})$ til $(E, \mathcal{E})$. Antag, at vi givet $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$ har en familie $(X_i)_{i \in I}$ af uafhængige stokastiske variable, hvor X_i har fordeling $P(\sigma_i, \cdot)$ på $(E, \mathcal{E})$, $i \in I$.*

Så er punktprocessen $\mu = \sum_{i \in I} \varepsilon_{(\sigma_i, X_i)}$ en poissonproces på $(S \times E, \mathbb{S} \otimes \mathcal{E})$ med intensitetsmål Λ, der er givet ved at for $A \in \mathbb{S}$ og $B \in \mathcal{E}$ er

$$\Lambda(A \times B) = \int_A P(s, B)\, \Phi(ds)$$

Bevis Vi skal bruge definition C.4.1 af en poissonproces, den tilhørende Laplacefunktional fra definition C.3.3 og sætning C.4.7. Det er klart, at μ er en simpel punktproces, da η er det, og vi som de begrænsede mængder kan bruge $A \times E$, hvor $A \in \mathbb{S}$ er en begrænset delmængde af S. At μ er en poissonproces med det angivne intensitetsmål, følger ved at betragte Laplacefunktionalen for μ. Lad derfor $f : S \times E \to \mathbb{R}_0$ være målelig og begrænset med kompakt støtte — igen kan mængder af formen $A \times E$ bruges, med $A \in \mathbb{S}$ kompakt. At f har kompakt støtte sikrer, at de uendelige produkter og summer nedenfor reelt kun er endelige. Definer nu $g : S \to \mathbb{R}_0$ ved $g(s) = \mathrm{E} \exp(-f(s, X)) = \int_E \exp(-f(s, x))\, P(s, dx)$, hvor X er en stokastisk variabel med fordeling $P(s, \cdot)$. Med notationen $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$ får vi nu:

$$L_\mu(f) = \mathrm{E}\, e^{-\mu(f)}$$

$$
\begin{aligned}
&= \mathrm{E}\,\mathrm{E}(e^{-\mu(f)}|\eta) \\
&= \mathrm{E}\,\mathrm{E}\Big(e^{-\sum_{i\in I} f(\sigma_i,X_i)}\;\Big|\;\eta\Big) \\
&= \mathrm{E}\prod_{i\in I} g(\sigma_i) && \text{pga. uafh.} \\
&= \mathrm{E}\,e^{-\eta(-\log g)} \\
&= \exp\Big(-\int_S (1-g(s))\;\Phi(ds)\Big) && \text{da } \eta \text{ er poisson} \\
&= \exp\Big(-\int_{S\times E}\Big(1-e^{-f(s,x)}\Big)\;P(s,dx)\Phi(ds)\Big)
\end{aligned}
$$

og det ønskede følger. □

2.4.5 *Bemærkning* Sætningen er kendt som "positioning dependent marking" i [Kar91], eksempel 1.28, opgave 1.11 og eksempel 2.24. Den omtales også i [DVJ88, p. 205-206].

2.4.6 Sætningen foretager to udvidelser af modellen i forhold til beskrivelsen i 2.4.1 og 2.4.2.

For det første tillades fødselsprocessen at være på et generelt rum S i stedet for $\mathbb{R}$. I det følgende vil fødselsprocessen $\eta = \sum_{i\in I}\sigma_i$ overalt være på kalendertidsaksen $\mathbb{R}$ som i 2.4.1. En mulig anvendelse af udvidelsen er dog omtalt i bemærkning 2.4.8.

For det andet er fordelingen af levetiden X_i nu en fordeling på et generelt rum E. Dette er en vigtig udvidelse. Vi tænker nu på X_i som beskrivelsen af et helt livsforløb. F.eks. kan E være et rum af stokastiske processer, hvor tidsaksen svarer til alder af individet og processens værdi til tid a beskriver individets tilstand i alder a. Som demonstreret i [ABGK93] er rummet af mærkede punktprocesser en fleksibel klasse af processer, der kunne anvendes.

2.4.7 Vi vil ikke diskutere om det er rimeligt at antage, at fødslerne følger en poissonproces. Dette er behandlet i artiklen [Bri86] og tilhørende diskussion.

Se dog sætning 3.9.5 og bemærkning 3.9.7, hvor vi viser en måde at analysere observationer på uden at bruge poissonantagelsen.

2.4.8 *Bemærkning* (Baggrundsvariable) Vi har ikke ovenfor sagt noget om, hvordan baggrundsvariable kan inkluderes i modellen. Inspireret af [Sim80] kan vi knytte baggrundsvariable til hvert individ samtidig med fødslerne. Fødselsprocessen η er nu på et rum $S = \mathbb{R}\times\tilde{S}$, hvor $s_1 \in \mathbb{R}$ angiver fødselstidspunktet og $s_2 \in \tilde{S}$ angiver værdien af baggrundsvariablene for individet født til tid s_1. Sætning 2.4.4 kan anvendes *to* gange. Først på $\mathbb{R}$ med en markovkerne fra $\mathbb{R}$ til $\tilde{S}$, hvor fødselstidspunkter og baggrundsvariable genereres. Den fremkomne poissonproces bruges nu som fødselsproces, hvor der til hver kombination af fødselstidspunkt og baggrundsvariable knyttes et livsforløb med en anden markovkerne.

2.4.9 *Bemærkning* (Intensitetsproces for Λ) Når Φ har intensitet β mht. Lebesguemålet l på $\mathbb{R}$, bliver intensitetsmålet Λ for μ målet $\beta(t)P(t,dx)dt$. Hvis yderligere alle målene $P(t,\cdot)$ har tæthed $g(t,x)$ mht. et mål π på E, så har Λ intensitet $\lambda(t,x)=\beta(t)g(t,x)$ mht. $l\otimes\pi$.

2.4.10 *Eksempel* Vi betragter nu det simple tilfælde, hvor $E=\mathbb{R}_0$ og X_i er levetiden for individet født til tid σ_i. Markovkernerne $P(\sigma,\cdot)$ har derfor tæthed $f_\sigma(a)=\exp(-\int_0^a\alpha(\sigma+u,u)\,du)\alpha(\sigma+a,a)$, svarende til beskrivelsen af levetiderne ved en todimensional hazard $\alpha(t,a)$ som i 2.3.1. Vi betragter modellen

$$\xrightarrow{\beta(t)} H \xrightarrow{\alpha(t,a)} D$$

Individer fødes ind i tilstanden H (healthy, rask) efter en poissonproces på kalendertidsaksen t med intensitet $\beta(t)$. Fra tilstand H er der to tidsakser — kalendertidsaksen t og aldersaksen a. Aldersaksen måler tiden siden individet startede i tilstand H. Tilstand H forlades til tilstand D (død) efter fordelingen med hazard $\alpha(t,a)$.

Resultaterne ovenfor giver direkte, at dødsfaldene i (kohorte,alder)-planen følger en poissonproces med intensitet

$$\tilde\lambda(t,a)=\beta(t)e^{-\int_0^a\alpha(t+u,u)\,du}\alpha(t+a,a)=\beta(t)f_t(a)$$

mht. l^2 for $t\in\mathbb{R}$ og $a\geqq 0$. Fortolkningen af $\tilde\lambda(t,a)$ er, at $\tilde\lambda(t,a)\,dt\,da$ er det forventede antal dødsfald i alder $]a,a+da]$ for individer født til tid $]t,t+dt]$.

Afbildningen $(t,a)\overset{f}{\mapsto}(t+a,a)$ flytter punkterne over i (kalendertid,alder)-planen; vores Lexis diagram. Her får vi ifølge korollar C.4.15 igen en poissonproces. Da f er en bijektiv lineær afbildning med determinant 1, er intensitetsprocessen på Lexis diagrammet

$$\lambda(t,a)=\beta(t-a)e^{-\int_0^a\alpha(t-a+u,u)\,du}\alpha(t,a)=\beta(t-a)f_{t-a}(a)$$

mht. $f(l^2)=l^2$. Tilsvarende er fortolkningen af $\lambda(t,a)$, at $\lambda(t,a)\,dt\,da$ er det forventede antal dødsfald i alder $]a,a+da]$ ved kalendertid $]t,t+dt]$.

Dette resultat findes i [Bri86].

2.5 Eksempel: invalidemodellen

2.5.1 Vi gennemregner i dette afsnit den invalidemodel, som er omtalt i [Kei91]. Formålet med dette er at illustrere styrken af sætning 2.4.4, samt nogle af de spørgsmål, den ikke giver svar på. Vi lægger vægt på de fordelingsmæssige udsagn om modellen, hvor [Kei91] hæfter sig ved sammenhængen prevalens = incidens × varighed, og estimation af invalideintensiteten ud fra et tværsnit til tid t.

2.5.2 Vi betragter invalidemodellen

$$\begin{tikzcd} {} \arrow[r, "\beta(t)"] & H \arrow[rr, "\gamma(t,a)"] \arrow[dr, "\alpha_H(t,a)"'] & & I \arrow[dl, "\alpha_I(t,a,d)"] \\ & & D & \end{tikzcd}$$

med de tre tilstande H (healthy, rask), I (inaktiv, invalid) og D (død).

Individer fødes ind i tilstand H efter en poissonproces på kalendertidsaksen t med intensitet $\beta(t)$. Fra tilstand H er der to tidsakser — kalendertidsaksen t og aldersaksen a. Aldersaksen måler tiden siden individet startede i tilstand H. De to intensiteter $\gamma(t,a)$ og $\alpha_H(t,a)$ styrer, hvornår individet forlader tilstand H, og om det sker til tilstanden D eller I. Hvis det sker til tilstand I, bestemmer intensiteten $\alpha_I(t,a,d)$ hvornår overgangen til tilstanden D finder sted. Det tredie tidsargument d i α_I er varigheden (duration) af opholdet i invalidetilstanden. Vi har en tredie tidsakse ud over kalendertid t og alder a.

2.5.3 Vi vælger rummet E til at være rummet af cadlag funktioner $X = (X(a))_{a \geqq 0}$, der er stykkevis konstante med de mulige tilstande $\tilde{E} = \{H, I, D\}$ og med kun endeligt mange spring i begrænsede tidsintervaller. Processen X beskriver tilstanden, individet er i til alder a.

Lad σ være fødselstidspunktet for et individ. Markovkernerne $P(\sigma, \cdot)$ er bestemt fra overgangsintensiteterne. Vi får umiddelbart, at

$$P_\sigma(\{X_0 = H\}) = 1$$
$$P_\sigma(\{X_a = H\}) = \exp\left(-\int_0^a (\alpha_H(\sigma+u,u) + \gamma(\sigma+u,u))\, du\right)$$

for $a \geqq 0$.

Lad nu $T_1 = \inf_s(X_s \neq H)$, med $\inf \emptyset = +\infty$. Dvs. T_1 er alderen, hvor tilstanden H forlades. Fordelingen af T_1 er altså fordelingen på $\mathbb{R}_0$ med hazardfunktion $\alpha_H(\sigma+a,a) + \gamma(\sigma+a,a) = (\alpha_H+\gamma)(\sigma+a,a)$, $a \geqq 0$. Nu er

$$P_\sigma(X_a = I | T_1 = a) = 1 - P(X_a = D | T_1 = a) = \frac{\gamma(\sigma+a,a)}{(\alpha_H+\gamma)(\sigma+a,a)}$$

Udtrykket er veldefineret ns. Dvs. $(\alpha_H+\gamma)(\sigma+T_1,T_1) \neq 0$ ns. i fordelingen af T_1.

Hvis det første spring sker ved alder $T_1 = a$ til tilstand I, så er længden af opholdet i I bestemt ved

$$\begin{aligned} P_\sigma(X_s = I | T_1 = a, X_a = I) &= \exp\left(-\int_0^{s-a} \alpha_I(\sigma+a+u, a+u, u)\, du\right) \\ &= \exp\left(-\int_a^s \alpha_I(\sigma+u, u, u-a)\, du\right) \end{aligned}$$

for $s \geqq a \geqq 0$.

Vi får nu f.eks. at

$$
\begin{aligned}
&P_\sigma(X_a = I) \\
&= \int_0^a P_\sigma(X_a = I, T_1 \in du)\, du \\
&= \int_0^a P_\sigma(T_1 \in du) P_\sigma(X_u = I | T_1 \in du) P_\sigma(X_a = I | T_u \in du, X_u = I)\, du \\
&= \int_0^a \Big[e^{\left(-\int_0^u (\alpha_H+\gamma)(\sigma+v,v)\, dv\right)} (\alpha_H + \gamma)(\sigma + u, u) \cdot \\
&\qquad \frac{\gamma(\sigma+u,u)}{(\alpha_H+\gamma)(\sigma+u,u)} e^{\left(-\int_u^a \alpha_I(\sigma+v,v,v-u)\, dv\right)} \Big]\, du \\
&= \int_0^a \Big[e^{\left(-\int_0^u (\alpha_H+\gamma)(\sigma+v,v)\, dv\right)} \gamma(\sigma+u,u) e^{\left(-\int_u^a \alpha_I(\sigma+v,v,v-u)\, dv\right)} \Big]\, du
\end{aligned}
$$

for $a \geqq 0$.

2.5.4 *Bemærkning* Processerne X beskrevet ovenfor bliver ikke markovprocesser i alderen a, da overgangssandsynlighederne fra tilstanden I afhænger af hvor længe vi har været i I; varigheden d.

2.5.5 (Parameterområde for α_H, α_I og γ) Funktionerne α_H, α_I og γ skal selvfølgelig være målelige. Yderligere vil vi kræve, at det for alle fødselstidspunkter σ er opfyldt, at $P_\sigma(\inf_s\{X_s = D\} < \infty) = 1$, med konventionen $\inf \emptyset = +\infty$. Dvs. vi ns. når tilstanden D i endelig tid. Dette er tilfældet, hvis følgende to betingelser er opfyldt:

1. For alle $\sigma \in \mathbb{R}$ er en nødvendig og tilstrækkelig betingelse for at vi forlader tilstand H i endelig tid, at $\int_0^\infty (\alpha_H + \gamma)(\sigma + u, u)\, du = \infty$.

2. For alle $a > 0$ og $\sigma \in \mathbb{R}$ skal $\int_a^\infty \alpha_I(\sigma + u, u, u - a)\, du = \infty$. Denne betingelse er tilstrækkelig og næsten nødvendig[2] for at vi kun tilbringer endelig lang tid i tilstand I.

Det vil endvidere være bekvemt, men ikke strengt nødvendigt, at kræve at levetiden, tiden fra fødsel til tilstand D nås, har endelig middelværdi. Et andet naturligt krav optræder i 2.5.15. Vi vender tilbage til parameterområdet i afsnit 3.2 i en situation uden invalidetilstanden, hvor vi interesserer os for estimation af α_H.

2.5.6 *Notation* $\mu = \sum_{i \in I} \varepsilon_{(\sigma_i, X_i)}$ er poissonprocessen af fødselstidspunkter og livsforløb på $\mathbb{R} \times E$.

[2] Det skal kun gælde for de aldre, hvor vi kan komme til tidstand I. Vi kan derfor sige at det kun skal gælde for γ-na a.

2.5.7 **Definition** (Nye processer ud fra μ) Definer først for $X \in E$

$$T_1 = \inf_s (X_s \neq H)$$
$$T_2 = \inf_s (X_s = D)$$

med konventionen $\inf \emptyset = +\infty$. T_1 er alderen for første spring. Hvis første spring sker fra H til D er $T_2 = T_1$, og hvis første spring sker til I, så er T_2 alderen for andet spring. Disse udsagn gælder næsten sikkert[3].

Lad $t \in \mathbb{R}$, $A, B \in \mathbb{B}_0$ og $C \in \mathbb{B}$. Definer følgende størrelser:

$$\begin{aligned}
N^{t,H}(A) &= \{\text{antal individer i tilstand } H \text{ til tid } t \text{ med alder} \in A \\
&= \mu((\sigma, X) : \sigma < t, X_{t-\sigma} = H, t - \sigma \in A) \\
N^{t,I}(A \times B) &= \begin{cases} \text{antal individer i tilstand } I \text{ til tid } t \text{ med alder} \in A \text{ og} \\ \text{varighed} \in B \end{cases} \\
&= \mu((\sigma, X) : \sigma < t, X_{t-\sigma} = I, t - \sigma \in A, (t - \sigma) - T_1 \in B) \\
N^{t,H,D}(C \times A) &= \begin{cases} \text{antal individer i tilstand } D \text{ til tid } t\text{, som kommer fra} \\ \text{tilstand } H \text{ og foretog overgangen } H \to D \text{ i kalendertid} \\ \in C \text{ og med alder } T_1 \in A \end{cases} \\
&= \mu((\sigma, X) : \sigma < t, X_{t-\sigma} = D, X_{T_1} = D, T_1 \in A, \sigma + T_1 \in C) \\
N^{t,I,D}(C \times A \times B) &= \begin{cases} \text{antal individer i tilstand } D \text{ til tid } t\text{, som kommer fra} \\ \text{tilstand } I \text{ og foretog overgangen } I \to D \text{ i kalendertid} \\ \in C \text{ og med alder } T_2 \in A \text{ og varighed } T_2 - T_1 \in B \end{cases} \\
&= \mu((\sigma, X) : \sigma < t, X_{t-\sigma} = D, X_{T_1} = I, T_2 \in A, \\
&\quad T_2 - T_1 \in B, \sigma + T_2 \in C) \\
N^{H,I}(C \times A) &= \begin{cases} \text{antal individer der i alder} \in A \text{ og kalendertid} \in C \\ \text{foretager overgangen } H \to I \end{cases} \\
&= \mu((\sigma, X) : T_1 \in A, \sigma + T_1 \in C, X_{T_1} = I)
\end{aligned}$$

2.5.8 **Sætning** (Uafhængige poissonprocesser) $N^{t,H}(\cdot)$, $N^{t,I}(\cdot)$, $N^{t,H,D}(\cdot)$, $N^{t,I,D}(\cdot)$ *og* $N^{H,I}(\cdot)$ *har versioner, der er poissonprocesser på henholdsvis aldersakse* $\mathbb{R}_0$*, aldersakse* $\times$ *varighed* $\mathbb{R}_0 \times \mathbb{R}_0$*, kalendertid* $\times$ *aldersakse* $\mathbb{R} \times \mathbb{R}_0$*, kalendertid* $\times$ *aldersakse* $\times$ *varighed* $\mathbb{R} \times \mathbb{R}_0 \times \mathbb{R}_0$ *og kalendertid* $\times$ *aldersakse* $\mathbb{R} \times \mathbb{R}_0$*.*

Disse versioner af processerne $N^{t,H}$*,* $N^{t,I}$*,* $N^{t,H,D}$ *og* $N^{t,I,D}$ *er uafhængige.*

Intensitetsprocesserne er

- *for* $N^{t,H}$

$$\lambda^{t,H}(a) = \beta(t-a) \exp\left(- \int_0^a (\alpha_H + \gamma)(t - a + u, u)\, du \right)$$

[3] Det betyder ikke noget hvordan vi definerer X_∞, da det alligevel er i en nulmængde det bliver aktuelt. Vi kunne f.eks. tilføje en ekstra tilstand Δ og definere $X_\infty = \Delta$.

- *for* $N^{t,I}$

$$\lambda^{t,I}(a,d) = 1\,(d \leqq a)\,\beta(t-a)\exp\left(-\int_0^{a-d}(\alpha_H+\gamma)(t-a+u,u)\,du\right)\cdot$$
$$\gamma(t-d,a-d)\exp\left(-\int_{a-d}^{a}\alpha_I(t-a+u,u,u-(a-d))\,du\right)$$

- *for* $N^{t,H,D}$

$$\lambda^{t,H,D}(s,a) = 1\,(s \leqq t)\,\beta(s-a)\exp\left(\int_0^{a}(\alpha_H+\gamma)(s-a+u,u)\,du\right)\alpha_H(s,a)$$

- *for* $N^{t,I,D}$

$$\lambda^{t,I,D}(s,a,d) = 1\,(s \leqq t, d \leqq a)\,\beta(s-a)\exp\left(-\int_0^{a-d}(\alpha_H+\gamma)(s-a+u,u)\,du\right)\cdot$$
$$\gamma(s-d,a-d)\exp\left(-\int_{a-d}^{a}\alpha_I(s-a+u,u,u-(a-d))\,du\right)\alpha_I(s,a,d)$$

- *for* $N^{H,I}$

$$\lambda^{H,I}(t,a) = \beta(t-a)\exp\left(-\int_0^{a}(\alpha_H+\gamma)(t-a+u,u)\,du\right)\gamma(t,a)$$

Bevis N processerne har versioner, der er poissonprocesser ifølge sætning C.4.12. (Rummet $\mathbb{R}\times E$ er et Borelrum, da vi kan skrive det som et følgerum. For begrænsede mængder i processerne vedrører hændelsen kun en begrænset mængde for fødselsprocessen η, og de forventede antal er derfor endelige. Etpunktsmængder vedrører etpunktsmængder for kontinuerte fordelinger, og det forventede antal er derfor 0. For $N^{t,H,D}$, $N^{t,I,D}$ og $N^{H,I}$ får vi ikke disjunkte mængder af $\mathbb{R}\times E$ for disjunkte mængder af argumentet[4]. Imidlertid har fællesmængderne intensitetsmål 0, da hvert individ næsten sikkert i alle fordelingerne $P_s(\cdot)$, $s\in\mathbb{R}$, kun foretager hver af overgangene 0 eller 1 gang.) Processerne $N^{t,H}$, $N^{t,I}$, $N^{t,H,D}$ og $N^{t,I,D}$ er endvidere uafhængige ifølge korollar til sætning C.4.12. (Den ikke tomme fællesmængde for $N^{t,H,D}(\mathbb{R}\times\mathbb{R}_0)$ og $N^{t,I,D}(\mathbb{R}\times\mathbb{R}_0\times\mathbb{R}_0)$ har intensitetsmål 0).

Vi viser udregningen af intensiteten for $N^{t,I}$. De resterende følger ved lignende regninger. For $A,B\in\mathbb{B}_0$ bemærker vi, at $t-s\in A$ er ækvivalent med $s\in t-A$ og tilsvarende $t-s-T_1\in B$ er ækvivalent med $T_1\in t-s-B$. Endvidere giver

[4] Hvis vi lader E være rummet af stykkevis konstante funktioner, der starter i tilstand H, derefter springer til tilstand D eller I, og hvis springet skete til I så springer til D, vil vi få rigtige disjunkte mængder. I dette tilfælde bliver N processerne selv poissonprocesser. Dette er en anvendelse af bemærkning C.4.13

kravet $t - s \in A$ automatisk at $s \leqq t$, og tilsvarende giver $t - s - T_1 \in B$ at $T_1 \leqq t - s$. Sætning 2.4.4 giver, at

$$
\begin{aligned}
&\mathrm{E}\, N^{t,I}(A \times B) \\
&= \int_{\mathbb{R}\times E} 1\,(s < t, t-s \in A, X_{t-s} = I, (t-s) - T_1 \in B)\; \beta(s)\, P_s(dx)\, ds \\
&= \int_{t-A} \beta(s) P_s(X_{t-s} = I, t - s - T_1 \in B)\, ds \\
&= \int_{t-A}\int_{t-s-B} \Big[\beta(s) \exp\Big(-\int_0^v (\alpha_H + \gamma)(s+u, u)\, du\Big) \cdot \\
&\qquad \gamma(s+v, v) \exp\Big(-\int_v^{t-s} \alpha_I(s+u, u, u-v)\, du\Big)\Big]\, dv\, ds \\
&= \int_{A\times B} \Big[1\,(\tilde{v} \leqq a)\, \beta(t-a) \exp\Big(-\int_0^{a-\tilde{v}} (\alpha_H + \gamma)(t-a+u, u)\, du\Big) \cdot \\
&\qquad \gamma(t-\tilde{v}, a-\tilde{v}) \exp\Big(-\int_{a-\tilde{v}}^{a} \alpha_I(t-a+u, u, u-(a-\tilde{v}))\, du\Big)\Big]\; d(a, \tilde{v})
\end{aligned}
$$

hvor vi bl.a. har udnyttet en regning som i 2.5.3 og variabelskiftet $(a, \tilde{v}) = (t - s, t - s - v)$, der giver $v = a - \tilde{v} \geqq 0$. □

2.5.9 *Bemærkning* I det følgende tænker vi altid på de versioner af N processerne, der er poissonprocesser.

2.5.10 *Bemærkning* Processerne $N^{t,H}$ og $N^{t,I}$ udtaler sig om fordelingen af individer på tilstande, mens processerne $N^{t,H,D}$ og $N^{t,I,D}$ udtaler sig om hvornår overgangen til tilstand D er sket.

2.5.11 *Bemærkning* Processen $N^{H,I}$ ("disease onset"), der angiver hvornår i Lexis diagrammet et individ overgår til invalidetilstanden, bliver ikke uafhængig af de andre processer, $N^{t,H}$, $N^{t,I}$, $N^{t,H,D}$ og $N^{t,I,D}$. Hvis vi kun kigger på overgange $H \to I$, der er sket før kalendertid t, så bliver processen uafhængig af $N^{t,H}$ og $N^{t,H,D}$.

2.5.12 *Bemærkning* Processerne $N^{t,H,D}$, $N^{t,I,D}$ og $N^{H,I}$ tæller antallet af *individer*, der foretager en bestemt overgang. Det er vigtigt, at hvert individ næsten sikkert kun foretager overgangen 0 eller 1 gang. Antallet af individer, der foretager overgangen, er derfor sammenfaldende med antal *overgange*, som foretages.

2.5.13 *Bemærkning* (Reaktivering) Overgangen $I \to H$ kalder vi reaktivering. Hvis vi tillader reaktivering får vi to typer af varighed i I. Vi kan både tale om varighed af det aktuelle ophold i I, og det akkumulerede tidsrum i I. Ses bort fra denne lille komplikation — og eventuelt ny definition af $N^{t,I}$ og $N^{t,I,D}$ — er det klart, at vi også med reaktivering får, at poissonprocesserne $N^{t,H}$, $N^{t,I}$, $N^{t,H,D}$ og $N^{t,I,D}$ er uafhængige. Intensiteterne vil ændre sig, og det er ikke nemt at skrive pæne udtryk op, da overgangssandsynlighederne for processerne X bliver mere indviklede.

Som $N^{H,I}$ er defineret bliver den ikke en simpel punktproces, når vi tillader reaktivering. Se bemærkning 2.6.10.

2.5.14 *Bemærkning* Vi har i beviset for sætning 2.5.8 regnet direkte på udtryk, der vedrører kalendertid og ikke fødselskohorte. Den translation vi udfører i eksempel 2.4.10, for at udskifte kohorten med kalendertid, sker i beviset ved, at vi ændrer integrationsvariabel i en eller flere integrationer undervejs.

2.5.15 Vi vil nu yderligere om parametrene β, α_H, γ og α_I kræve, at

$$\begin{aligned} \mathrm{E}\, N^{t,H} &= \int_0^\infty \lambda^{t,H}(a)\, da < \infty \\ \mathrm{E}\, N^{t,I} &= \int_{\mathbb{R}_0^2} \lambda^{t,I}(a,v)\, d(a,v) < \infty \end{aligned}$$

hvor $\mathrm{E}\, N^{t,H}$ og $\mathrm{E}\, N^{t,I}$ er det forventede antal individer i tilstand H hhv. I til tidspunkt t. Se afsnit 3.2 for tilstrækkelige betingelser i modellen uden invalidetilstanden.

Resultaterne i sætning 2.5.8 giver direkte, at antallet af personer $N^{t,H} = N^{t,H}(\mathbb{R}_0)$ og $N^{t,I} = N^{t,I}(\mathbb{R}_0 \times \mathbb{R}_0)$ i tilstand H og I til tidspunkt t er stokastisk uafhængige og poissonfordelte.

Bemærkning C.4.5 giver, at hvis vi betinger med antallet af personer i live til tid t, $N^{t,H} + N^{t,I} = n$, og derefter trækker et tilfældigt af de n individer, så er fordelingen af status, alder og varighed givet ved

$$\begin{aligned} &P(\text{individ i } H) = 1 - P(\text{individ i } I) = \frac{\mathrm{E}\, N^{t,H}}{\mathrm{E}\, N^{t,H} + \mathrm{E}\, N^{t,I}} \\ &P(\text{alder } \in da | \text{individ i } H) = \frac{\lambda^{t,H}(a)}{\mathrm{E}\, N^{t,H}}\, da \\ &P(\text{alder } \in da, \text{varighed } \in dv | \text{individ i } I) = \frac{\lambda^{t,I}(a,v)}{\mathrm{E}\, N^{t,I}}\, da\, dv \end{aligned}$$

2.5.16 Én version af formlen

$$\text{prævalens} = \text{incidens} \times \text{varighed}$$

er $\mathrm{E}\, N^{t,H} = \beta M$. Den fås når vi har kalendertidshomogenitet, dvs. β, α_H, γ og α_I ikke afhænger af kalendertiden t. Da er

$$\mathrm{E}\, N^{t,H} = \beta \int_0^\infty \exp\left(- \int_0^a (\alpha_H + \gamma)(u)\, du \right) da = \beta M$$

hvor M er middellevetiden i tilstand H for et nyfødt individ.

2.5.17 *Eksempel* (Den simple model $\xrightarrow{\beta(t)} H \xrightarrow{\alpha(t,a)} D$) Som ovenfor får vi i det simple eksempel 2.4.10, at for $A \in \mathbb{B}_0$ er

$$\begin{aligned} N^t(A) &= \begin{cases} \text{antal individer i tilstand } H \text{ til kalendertid } t \text{ med alder} \\ \in A \end{cases} \\ &= \mu((\sigma, X) : \sigma < t, X > t - \sigma, t - \sigma \in A) \end{aligned}$$

en poissonproces med intensitet

$$\lambda^t(a) = \beta(t-a) \exp\left(-\int_0^a \alpha(t-a+u, u)\, du\right)$$

Eksempel 2.4.10 siger, at for $A \in \mathbb{B}_0$ og $C \in \mathbb{B}$ er

$$\begin{aligned} N^{H,D}(C \times A) &= \begin{cases} \text{antal individer, der i alder } \in A \text{ og kalendertid } \in C \\ \text{foretager overgangen } H \to D \end{cases} \\ &= \mu((\sigma, X) : \sigma + X \in C, X \in A) \end{aligned}$$

en poissonproces med intensitet

$$\lambda(t,a) = \beta(t-a) e^{-\int_0^a \alpha(t-a+u,u)\, du} \alpha(t,a) = \beta(t-a) f_{t-a}(a)$$

En bemærkning som 2.5.11 er naturligvis også gældende her.

2.5.18 Den generelle sætning 2.4.4 er i dette afsnit anvendt på invalidemodellen. Vi har fået mange konkrete resultater om uafhængige poissonprocesser i invalidemodellen, sætning 2.5.8.

2.6 Struktur af livsforløb

2.6.1 I princippet er alle resultater indeholdt i sætning 2.4.4, men som det foregående afsnit viser, så kan vi i konkrete situationer nemt få nye meget anvendelige resultater. Vi har med andre ord en balancegang mellem at skrive generelle resultater op, sætning 2.4.4, og at anvende dem i konkrete situationer, sætning 2.5.8. I dette afsnit prøver vi at finde en balancegang mellem de to situationer.

2.6.2 Lad $(\tilde{E}, \tilde{\mathcal{E}})$ være et rum af tilstande et individ kan befinde sig i. Som rummet E af livsforløb vil vi bruge de stykkevis konstante cadlag[5] processer med tilstandsrum $\tilde{E}$, der kun har endeligt mange spring på begrænsede intervaller. For $X = (X_a)_{a \geqq 0} \in E$ er X_a den tilstand individet er i til alder a, $a \geqq 0$.

Rummet E kan også beskrives som en mærket punktproces, [Jac95a, eks. 2.13].

[5] Vi skal f.eks. have en topologi på $\tilde{E}$ for, at vi kan tale om højrekontinuitet.

2.6.3 I invalidemodellen i afsnit 2.5 fik vi ud over tidsakserne alder og kalendertid også en tidsakse, vi kaldte varighed af invaliditet. I den ramme, vi er i nu, får vi langt flere mulige ekstra tidsakser. Uden at definere disse præcist kan vi sige, at en tidsakse er aldersafstanden mellem forskellige typer af spring for processen X.

2.6.4 **Definition** Definer for $t \in \mathbb{R}$, $B \in \tilde{\mathcal{E}}$ og $A \in \mathbb{B}_0$

$$\begin{aligned} N^{t,B}(A) &= \begin{cases} \text{antal individer til tid } t \text{ i en tilstand } \in B \text{ med alder} \\ \in A \end{cases} \\ &= \mu((\sigma, X) : \sigma < t, X_{t-\sigma} \in B, t - \sigma \in A) \end{aligned}$$

2.6.5 **Sætning** *For $t \in \mathbb{R}$ og $B_1, \ldots, B_n \in \tilde{\mathcal{E}}$ parvis disjunkte har $N^{t,B_1}, \ldots, N^{t,B_n}$ versioner, der er uafhængige poissonprocesser på aldersaksen $\mathbb{R}_0$. Intensitetsprocessen for N^{t,B_i}, $0 \leqq i \leqq n$, er*

$$\lambda^{t,B_i}(a) = \beta(t-a) P_{t-a}(X_a \in B_i)$$

Bevis er mage til beviset for sætning 2.5.8. □

2.6.6 *Bemærkning* Sætningen er om, hvilke tilstande individerne befinder i til tidspunkt t. Vi har ikke forsøgt at inddrage alle de mulige ekstra varighedsakser. Disse må inddrages i en konkret anvendelse.

2.6.7 *Bemærkning* Hvis vi har en tilstand $D \in \tilde{E}$, død, hvor alle ender, så vil der naturligvis være uendeligt mange personer i denne tilstand. Her lader vi imidlertid aldersaksen fortsætte efter, at personen er død. I sætning 2.5.8 registrerede vi alder ved overgang til tilstanden D. Vi inddrog samtidig kalendertiden for overgangen, for at få endeligt mange punkter i begrænsede mængder.

2.6.8 **Definition** Definer for $B_1, B_2 \in \tilde{\mathcal{E}}$ disjunkte, $C \in \mathbb{B}$ og $A \in \mathbb{B}_0$:

$$\begin{aligned} N^{B_1,B_2}(C \times A) &= \begin{cases} \text{antal individer der i kalendertid } \in C \text{ og alder } \in A \\ \text{foretager et spring fra en tilstand } \in B_1 \text{ til en tilstand} \\ \in B_2 \end{cases} \\ &= \mu((\sigma, X) : \exists t \in C \text{ så } t - \sigma \in A, X_{(t-\sigma)-} \in B_1, X_{t-\sigma} \in B_2) \end{aligned}$$

2.6.9 **Sætning** *Lad $B_1, B_2 \in \tilde{\mathcal{E}}$ være disjunkte. Antag, at for Φ-næsten alle $s \in \mathbb{R}$ gælder*

$$P_s(\exists a_1, a_2 : 0 \leqq a_1 < a_2, X_{a_1} \in B_2, X_{a_2} \in B_1) = 0$$

(For Λ-na individer sker overgang fra en tilstand i B_2 til en tilstand i B_1 ikke.)

Da har N^{B_1,B_2} en version, der er en poissonproces på kalendertidsakse × aldersakse $\mathbb{R} \times \mathbb{R}_0$.

Bevis følger som i beviset for sætning 2.5.8. □

2.6.10 *Bemærkning* Det er altafgørende, at antallet af forskellige individer, der foretager en overgang fra B_1 til B_2, er sammenfaldende med antallet af overgange fra B_1 til B_2, der finder sted. Hvis dette ikke var tilfældet, ville N^{B_1,B_2} overhovedet ikke blive et tællemål på kalendertidsakse $\times$ aldersakse, da vi ikke ville have additivitet for disjunkte delmængder.

Hvis vi i stedet i definition 2.6.8 tæller antal overgange fra B_1 til B_2 i kalendertid C og alder A, vil N^{B_1,B_2} blive et tællemål, men ikke generelt en poissonproces. Vi kan ikke vise uafhængighed for disjunkte mængder, da vi kun kan bygge på uafhængigheden for disjunkte mængder af $\mathbb{R} \times E$, som vi ikke kan føre det tilbage til. Man kan nemt forestille sig situationer, hvor antal overgange ikke er poissonfordelt. F.eks. hvis der med sandsynlighed 1 foretages to overgange fra B_1 til B_2 før alder 10 år. N^{B_1,B_2} vil dog give os tælleprocesser, som man kan studere i konkrete tilfælde.

2.6.11 *Bemærkning* Intensitetsprocessen bliver ikke pæn. Hvis T_n betegner det n'te springtidspunkt for processen X er

$$\begin{aligned}
&\mathrm{E}\, N^{B_1,B_2}(C \times A) \\
&\quad = \int_{\mathbb{R}} \beta(s) P_s(\cup_{t\in C}(t-s \in A, X_{(t-s)-} \in B_1, X_{t-s} \in B_2))\, ds \\
&\quad = \int_{\mathbb{R}} \beta(s) \sum_{n=1}^{\infty} P_s(s+T_n \in C, T_n \in A, X_{T_n-} \in B_1, X_{T_n} \in B_2)\, ds
\end{aligned}$$

2.6.12 *Bemærkning* Sætning udtaler sig om, hvilke overgange individerne foretager. Uden yderligere struktur på processerne, kan vi ikke opnå pæne resultater om uafhængighed mellem processer N^{B_1,B_2} for B_1, B_2 varierende, da det samme individ kan foretage flere overgange i et kalendertidsrum C og aldersrum A.

2.7 Kommentarer til litteraturen

2.7.1 Jeg har fundet fire artikler, der lader fødselstidspunkterene følge en poissonproces, og individer have uafhængige livsforløb givet fødselstidspunkterne. De fire artikler er [Sim80], [Bri86], [Kei91] og [Alh92]. Ifølge [Kei91, p. 372] er [Sim80] og [Bri86] de to eneste tidligere referencer, der benytter antagelsen om, at fødslerne følger en poissonproces på Lexis diagrammet.

Fælles for de fire artikler er, at de ikke lader individerne have et generelt livsforløb, men typisk kigger på levetiden for individerne, som i eksempel 2.4.10.

2.7.2 Richard Simon beskæftiger sig i [Sim80] med tiden siden diagnose for sygdom til død. Han antager, at diagnosetidspunkterne følger en homogen poissonproces, og udvælger personer ved at lave et tværsnit til kalendertid t og have en opfølgningsperiode Δ. Vi omtaler Simons udvælgelsesmetode i afsnit 3.6.

Simon interesserer sig for den situation, at individerne har en af to mulige levetidsfordelinger, der er bestemt af en observerbar binær baggrundsvariabel, som ikke er tidsafhængig. Endvidere skal levetiderne ikke afhænge af kalendertiden. De to levetidsfordelinger antages at stamme fra en en-parameter eksponential familie. Herefter opstiller Simon den relevante likelihoodfunktion og finder ligninger, der kan bruges til at estimere parametrene og finde deres asymptotiske fordeling.

2.7.3 I rammen af Lexis diagrammet, som beskrevet i eksempel 2.4.10, betragter Brillinger i [Bri86] visse "vital statistics" fra befolkningsteori. Han beskæftiger sig bl.a. med fordelingen af "dødsrater", antal døde/antal under observation, og kommer med grænser for disse.

I artiklen og en del af de tilhørende diskussioner drøftes det rimelige i antagelsen om, at fødslerne følger en poissonproces. Jeg vil ikke gå ind i denne diskussion i dette speciale.

2.7.4 I artiklen [Kei91] uddyber Keiding sin diskussion [Kei86] til Brillingers artikel. Keiding interesser sig for relationen

$$\text{prævalens} = \text{incidens} \times \text{varighed}$$

Han benytter rammen af invalidemodellen, som vi har gennemregnet i afsnit 2.5, til at udlede relationen i tilfælde af tidshomogenitet og for overgangen fra rask til invalid.

I anden halvdel af artiklen gennemgås, under forskellige forudsætninger, estimation af specielt intensiteten γ for invaliditet, fra et tværsnit til et fast kalendertidspunkt t.

2.7.5 Juha M. Alho udvider i [Alh92] relationen

$$\text{prævalens} = \text{incidens} \times \text{varighed}$$

fra det tidshomogene tilfælde i [Kei91] til at omfatte tilfældet, hvor populationen udvikler sig eksponentielt.

2.8 Konklusion

2.8.1 I dette kapitel har vi indført Lexis diagrammet. Lexis diagrammet, figur 2.1, beskriver livsforløb i (kalendertid,alder)-planen ved skrå linier med hældning 1.

2.8.2 Vi antager, at fødselstidspunkterne for individer følger en poissonproces, $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$, og at livsforløbene X_i, $i \in I$, for individer født til tidspunkt σ_i, $i \in I$, er uafhængige givet fødselsprocessen η.

Sætning 2.4.4 siger, at hvis vi kombinerer dette i $\mu = \sum_{i \in I} \varepsilon_{(\sigma_i, X_i)}$, så bliver μ en poissonproces på (kohorte,livsforløb)-planen. Den simultane fordeling af fødselstidspunkter og livsforløb er således beskrevet ved en poissonproces.

2.8.3 Invalidemodellen uden reaktivering er blevet gennemregnet i stor detalje i sætning 2.5.8. Den generelle sætning giver i dette specialtilfælde en række resultater om uafhængighed og fordeling af individerne på de tre tilstande.

2.8.4 I afsnit 2.6 forsøger vi at finde en balancegang mellem den generelle model og anvendelse i et konkret tilfælde. Vi tænker os hvert individs livsforløb repræsenteret ved en stykkevis konstant proces. Vi viser, at til et fast tidspunkt t er antallet af individer i disjunkte tilstande uafhængige poissonprocesser på aldersaksen. Vi kan ikke uden yderligere antagelser vise generelle sætninger om, hvordan (kalendertid,alder)-punkterne for overgange mellem tilstande fordeler sig. De yderligere antagelser er, at hvert individ kun foretager den interessante overgang 0 eller 1 gang.

2.8.5 Denne ramme at behandle Lexis diagrammet i, giver nemme beviser for fordelingsmæssige resultater.

Kapitel 3

Sampling i Lexis diagrammet

3.1 Introduktion

3.1.1 Vi interesserer os for sampling i den simple model $\stackrel{\beta(t)}{\longrightarrow} H \stackrel{\alpha(t,a)}{\longrightarrow} D$, som er beskrevet i eksempel 2.4.10 og 2.5.17. Personernes fødselstidspunkter er givet ved en poissonproces $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$ med intensitet $\beta(t)$. Givet η er levetiderne $(X_i)_{i \in I}$ uafhængige, og levetiden X_i for individet født til tid σ_i har tæthed $f_{\sigma_i}(x) = \exp\left(-\int_0^x \alpha(\sigma_i + a, a)\, da\right) \alpha(\sigma_i + x, x)$. Den tilhørende overlevelsesfunktion kaldes $\overline{F}_{\sigma_i}(x)$. I tilfælde af tidshomogenitet bruges betegnelserne $f(x)$, $\overline{F}(x)$ og $\alpha(x)$ for tæthed, overlevelsesfunktion og hazard.

3.1.2 Vi vil i vores model undersøge forskellige måder at udvælge personer på. Ved at opstille likelihoodfunktionen, vil vi undersøge i hvilke fordelinger, vi skal finde en ikke-parametrisk maksimum likelihood estimator for levetidsfordelingen.

Vi får de interessanteste resultater i det tidshomogene tilfælde, hvor intensiteterne $\beta(t) = \beta$ og $\alpha(t, a) = \alpha(a)$ ikke afhænger af kalendertiden t. Generelt vil vi ikke interessere os for estimation af β.

Specielt vil vi undersøge, under hvilke forudsætninger vi kan bruge de sædvanlige estimationsteknikker fra tælleprocesser med venstretrunkering/højrecensur og Nelson-Aalen estimatoren. Det viser sig, at hvis vi betinger med fødselstidspunkterne for de individer vi ser, så får vi denne analyse. Dette resultat afhænger ikke af, at fødselstidspunkterne følger en poissonproces.

3.1.3 I oversigtsartiklen [Kei90] er der en gennemgang af forskellige forslag til udvidelse af den sædvanlige tælleprocesteori til to tidsargumenter, herunder estimation af den underliggende parameter, $\alpha(t, a)$, i det ikke tidshomogene tilfælde. Vi beskæftiger os ikke med disse vigtige problemer.

3.1.4 Vi har i kapitel 2 opstillet en model for fødsler og livsforløb.

Vi vil i denne model gennemgå fem typer af udvælgelse af personerne:

- Vi laver et tværsnit til kalendertidspunkt t, og ser alle personer i live til kalendertidspunkt t.
- Vi observerer alle dødsfald i et kalendertidsinterval.
- Vi ser alle personer i live og alle dødsfald i et kalendertidsinterval.
- Vi ser et tværsnit til tid t_1 og følger personer i live ved tid t_1 til tid t_2. Dette er rammen fra [Sim80].
- Vi følger en kohorte — individer født i en kalendertidsperiode.

Vi vil regne på disse fem typer af observation under forskellige forudsætninger for fødslerne:

- Fødslerne følger en poissonproces som i kapitel 2.
- Vi betinger med nogle af fødselstidspunkterne. Typisk fødselstidspunkter for individer vi ser i live.

3.1.5 Vi giver i afsnit 3.2 nogle overvejelser om hvilket parameterområde, vores intensiteter ligger i.

De efterfølgende afsnit 3.3–3.7 gennemgår de fem udvælgelsesmønstre ovenfor, når fødslerne følger en poissonproces. I afsnit 3.8 behandler vi de tre typer af elementærmængder for død i Lexis diagrammet.

I afsnit 3.9 behandler vi generelt hvad der sker, når vi betinger med visse fødselstidspunkter. Afsnit 3.10 og 3.11 omtaler begge et tværsnit til tid t, hvor vi har betinget med alle fødselstidspunkter før tid t hhv. fødselstidspunkterne for de individer vi ser. Afsnit 3.12 behandler i en generel ramme de resterende tilfælde, og når vi betinger med fødselstidspunkterne, kan vi uden problemer analysere aldrene vha. tælleprocesser.

Resultaterne er resumeret i afsnit 3.13.

3.1.6 Vi vil i starten af kapitlet overveje detaljer, som senere kunne gentages i meget lignende situationer, men ikke bliver det.

3.2 Parameterområde

3.2.1 De eneste krav, vi har stillet til intensiteten $\beta(t)$ for fødselsprocessen og hazarden $\alpha(t,a)$ for levetiderne er, at

- De skal være målelige funktioner.
- For begrænsede intervaller A skal $\int_A \beta(t)\,dt < \infty$, hvilket sikrer, at der kun fødes endeligt mange individer i begrænsede intervaller.

- For fast fødselstidspunkt $\sigma \in \mathbb{R}$ skal $\exp\left(-\int_0^x \alpha(\sigma+a,a)\,da\right)\alpha(\sigma+x,x)$, $x \in \mathbb{R}_0$, være en tæthed på $\mathbb{R}_0$, der beskriver levetiden for individet født til tid σ. Dette er tilfældet hvis og kun hvis $\int_0^\infty \alpha(\sigma+a,a)\,da = \infty$.

3.2.2 *Notation* Lad N^t være poissonprocessen af aldre ved et tværsnit til tid t. Denne er omtalt i eksempel 2.5.17.

Sæt $N^t = N^t(\mathbb{R}_0)$, antal personer i live til tid t. Denne misbrug af notation burde ikke give problemer i det følgende.

3.2.3 I det følgende afsnit 3.3 observerer vi alle personer i live til et kalendertidspunkt t og deres aldre. Ifølge eksempel 2.5.17 er aldrene for individerne i live til tid t en poissonproces N^t med intensitet

$$\lambda^t(a) = \beta(t-a)\exp\left(-\int_0^a \alpha(t-a+u,u)\,du\right)$$

Det vil være naturligt at kræve, at det forventede antal personer i live til et fast kalendertidspunkt t er endeligt. Sætning C.4.8 giver i det tilfælde en likelihoodfunktion for vores observation. Med $N^t = N^t(\mathbb{R}_0)$ er det forventede antal individer

$$\mathrm{E}\,N^t = \int_0^\infty \lambda^t(a)\,da = \int_0^\infty \beta(t-a)\exp\left(-\int_0^a \alpha(t-a+u,u)\,du\right)da \quad (3.1)$$

Denne størrelse er ikke nødvendigvis endelig. Problemet opstår, fordi der er tale om individer født over et ubegrænset interval, $]-\infty,t]$. Der er to forhold at tage i betragtning:

For det første kan vi forestille os, at $\beta(t-a)$ holder sig begrænset, men at levetidsfordelingerne får tungere og tungere haler for $t-a \to -\infty$.

For det andet kan halen i levetidsfordelingen holde sig begrænset, men der bliver født flere og flere individer, jo længere tilbage i kalendertid $t-a$ ligger. Dvs. at $\beta(t-a)$ vokser for hurtigt for $a \to \infty$.

Vi vil i det følgende kræve, at

- $\mathrm{E}\,N^t < \infty$

3.2.4 I det tidshomogene tilfælde $\beta(t) = \beta$ og $\alpha(t,a) = \alpha(a)$ reducerer (3.1) til βM, hvor M er middellevetiden bestemt af intensiteten α. Dette motiverer kravet

- For alle $\sigma \in \mathbb{R}$ skal middelværdien M_σ i fordelingen med tæthed $f_\sigma(x)$, $x \in \mathbb{R}_0$, være endelig.

Kravet synes rimeligt og ikke særlig restriktivt.

3.2.5 **Definition** Om to reelle stokastiske variable X og Y siger vi, at X er stokastisk mindre end Y, hvis

$$\forall v \in \mathbb{R} : P(X \leqq v) \ \geqq \ P(Y \leqq v)$$

og vi skriver $X \overset{\mathcal{D}}{<} Y$.

3.2.6 *Bemærkning* Betingelsen for $X \overset{\mathcal{D}}{<} Y$ er selvfølgelig ækvivalent med

$$\forall v \in \mathbb{R} : P(X > v) \ \leqq \ P(Y > v)$$

3.2.7 *Notation* Lad i det følgende U_s være en stokastisk variabel, der følger fordelingen med tæthed f_s — levetidsfordelingen for individet født til tidspunkt s.

3.2.8 **Sætning** (Tilstrækkelige betingelser) *Følgende betingelser er hver for sig tilstrækkelige for, at* $\mathrm{E}\,N^t$ *er endelig,* $\mathrm{E}\,N^t < \infty$.

1. *For alle* $s < t$ *er* U_s *næsten sikkert begrænset af en konstant* C. *Dvs.* $U_s < C$ *ns.*

2. *For alle* $s < t$ *er* $\beta(s)$ *begrænset af en konstant* C *og der findes en stokastisk variabel* U, *hvis fordeling må afhænge af* α, *med endelig middelværdi, så* $U_s \overset{\mathcal{D}}{<} U$.

Bevis

1. Her er $\exp\left(-\int_0^a \alpha(t-a+u,u)\,du\right) = 0$ for $a > C$, så $\mathrm{E}\,N^t < \int_0^C \beta(t-a)\,da < \infty$, da β er intensiteten i en poissonproces.

2. Overlevelsesfunktionen for U kaldes $\overline{F}$.
Vi får $\mathrm{E}\,N^t \leqq \int_0^\infty \beta(t-a)\overline{F}(a)\,da \leqq CM$, hvor M er middellevetiden i fordelingen af U. (Hvis U følger en diskret fordeling er M endelig, men ikke nødvendigvis middelværdien.) $\square$

3.2.9 *Bemærkning* Sætningen giver tilstrækkelige kriterier for, at *et* bestemt par af funktioner $\beta(t)$ og $\alpha(t,a)$ gør middelværdien $\mathrm{E}\,N^t$ endelig.

Vi kan ikke give en pæn karakteristik af *familien* af par (β,α), som gør $\mathrm{E}\,N^t$ endelig.

3.2.10 *Bemærkning* Den anden betingelse er opfyldt i det tidshomogene tilfælde $\beta(t) = \beta$ og $\alpha(t,a) = \alpha(a)$.

3.2.11 *Eksempel* (Eksponentialfordelte overlevelsestider) Hvis β udvikler sig eksponentielt, $\beta(t) = C_1\exp(-C_2 t)$ med $C_1, C_2 > 0$, viser en direkte udregning, at hele familien af eksponentialfordelte levetider $\alpha(t,a) = \alpha$ med $\alpha > C_2$ opfylder

$\mathrm{E}\, N^t < \infty$. Vi har således her en β-funktion, der er eksponentielt voksende for $t \to -\infty$.

Hvis vi vil have alle eksponentialfordelingerne $\alpha(t,a) = \alpha$, $\alpha > 0$, med, kan vi ikke længere tillade, at β har eksponentiel form. Hvis β er begrænset, kan sætning 3.2.8 bruges. Direkte udregning viser, at f.eks. en potensfunktion $\beta(t) = C_1|t|^{C_2}$ kan bruges. Thi

$$\int_0^\infty \beta(t-a)\exp(-\alpha a)\, da = \int_0^\infty \underbrace{C_1|t-a|^{C_2} e^{-\frac{\alpha}{2}a}}_{*} \underbrace{e^{-\frac{\alpha}{2}a}}_{**}\, da$$

hvor $*$ går mod 0 for $a \to \infty$ og $**$ er integrabel. Dvs. integralet er endeligt.

3.3 Tværsnit

3.3.1 Vi vil i dette afsnit gennemgå estimation af α i det tilfælde, hvor vi observerer et tværsnit til kalendertid t. En "cross sectional" undersøgelse.

3.3.2 Til et fast kalendertidspunkt t observerer vi alderen for personer i live til dette kalendertidspunkt. Dette er illustreret i figur 3.1, hvor vi ser i hvilke aldre den lodrette stiplede streg ved tid t krydser en livslinie.

Vi har i eksempel 2.5.17 set, at disse aldre giver anledning til en poissonproces N^t med intensitet

$$\lambda^t(a) = \beta(t-a)\exp\left(-\int_0^a \alpha(t-a+u,u)\, du\right) = \beta(t-a)\overline{F}_{t-a}(a)$$

Vi har i afsnit 3.2 diskuteret betingelser for at $\mathrm{E}\, N^t = \int_0^\infty \lambda^t(a)\, da < \infty$. I det efterfølgende antager vi, at integralet er endeligt.

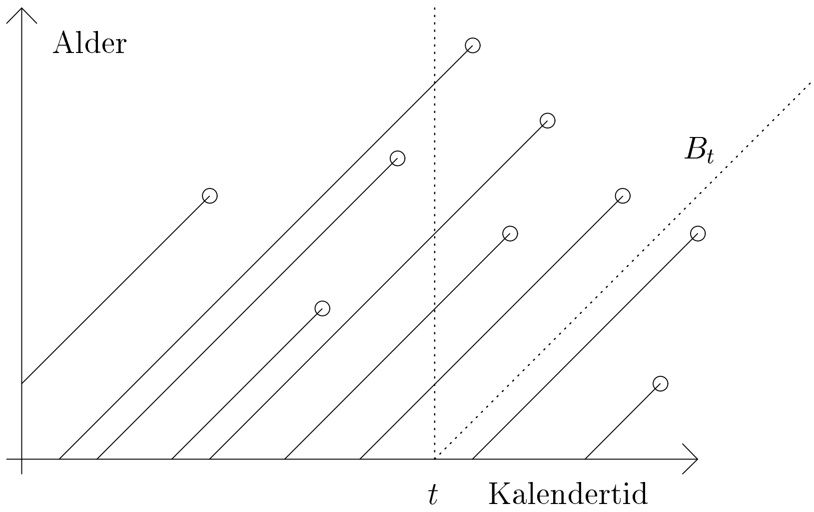
3.3.3 I eksempel 2.5.17 beskrev vi de aldre, vi ser ved et tværsnit, direkte vha. poissonprocessen μ. Vi kan også tage udgangspunkt i eksempel 2.4.10, hvor vi så, at dødsfaldene i Lexis diagrammet udgør en poissonproces med intensitet

$$\lambda(t,a) = \beta(t-a)e^{-\int_0^a \alpha(t-a+u,u)\, du}\alpha(t,a) = \beta(t-a)f_{t-a}(a)$$

Vi observerer en transformation af vores poissonproces af dødsfald i Lexis diagrammet. Dødsfaldene i trekanten

$$B_t = \{(s,a) \in \mathbb{R} \times \mathbb{R}_0 \,|\, s \geqq t, a \geqq s-t\}$$

transformeres med afbildningen $h : B_t \to \mathbb{R}_0$ givet ved $h(s,a) = a - (s-t)$. $h(s,a)$ er alderen til tid t for individet, der dør til tid s i alder a. Trekanten B_t er markeret med stiplede linier i figur 3.1.



Figur 3.1: Lexis diagram med illustration af tværsnit til tid t. Vi observerer i hvilke aldre livsforløbene skærer den lodrette stiplede linie til tid t. Trekanten B_t markeret med stiplede linier omtales i 3.3.3.

Ifølge korollar C.4.15 bliver den transformerede proces igen en poissonproces. Vi får selvfølgelig den samme poissonproces som i eksempel 2.5.17, hvilket følgende udregning viser:

$$
\begin{aligned}
h(\Lambda\,|_{B_t})([0,a]) &= \int_t^\infty \int_{s-t}^{s-t+a} \lambda(s,u)\,du\,ds \\
&= \int_0^a \int_t^\infty \lambda(s,\tilde{u}+(s-t))\,ds\,d\tilde{u} \\
&= \int_0^a \beta(t-\tilde{u}) \int_t^\infty f_{t-\tilde{u}}(\tilde{u}+(s-t))\,ds\,d\tilde{u} \\
&= \int_0^a \beta(t-\tilde{u}) \exp\left(-\int_0^{\tilde{u}} \alpha(t-\tilde{u}+x,x)\,dx\right) d\tilde{u}
\end{aligned}
$$

hvor variabelskiftet $(s,\tilde{u}) = (s, u-(s-t))$ er benyttet.

3.3.4 *Notation* Vi observerer poissonprocessen $N^t(\cdot)$ over hele aksen $\mathbb{R}_0$. Lad $a = (a_1, \ldots, a_{N^t})$ være de aldre vi observerer ved et tværsnit til tidspunkt t. Punkterne er nummereret tilfældigt, som omtalt i bemærkning C.4.5.

3.3.5 Likelihoodfunktionen for observationen er givet i sætning C.4.8 med tilhørende korollar. Som referencemål benytter vi poissonprocessen med samme form af intensiteten og parameterværdierne $\beta \equiv 1$, $\alpha \equiv 1$. Under referencemålet er det forventede antal personer i live til tid t også endeligt, $\int_0^\infty e^{-a}\,da = 1 < \infty$. Vi får likelihoodfunktionen

$$L = \exp\left(-\operatorname{E} N^t + 1\right) \prod_{i=1}^{N^t} \frac{\lambda^t(a_i)}{\exp(-a_i)}$$

I det tidshomogene tilfælde, hvor hverken β eller α afhænger af kalendertiden, reducerer dette til

$$L \propto \exp\left(-\beta M\right) \beta^{N^t} \prod_{i=1}^{N^t} \exp\left(-\int_0^{a_i} \alpha(u)\, du\right) \tag{3.2}$$

hvor $M = \int_0^\infty \exp(-\int_0^a \alpha(u)\, du)\, da$ er middelværdien i fordelingen bestemt af α.

Det er ikke umiddelbart klart, hvordan vi skal estimere i denne likelihoodfunktion.

3.3.6 *Bemærkning* For at få f.eks. (3.2) er det selvfølgelig tilstrækkeligt at kræve tidshomogenitet af β for fødsler, der er involveret i mængden, der observeres over, samt tidshomogenitet af intensiteten for den relevante mængde. F.eks. skal kun $\beta(s)$ og $\alpha(s,a)$ for $s \leqq t$ ikke afhænge af s.

3.3.7 Hvis vi betinger med $N^t = n$, så er $a_1, \ldots, a_n$ ifølge bemærkning C.4.5 iid med fordeling med tæthed

$$\frac{\lambda^t(a)}{\operatorname{E} N^t} = \frac{\beta \exp\left(-\int_0^a \alpha(u)\, du\right)}{\beta M} = \frac{\exp\left(-\int_0^a \alpha(u)\, du\right)}{M}$$

Når vi betinger med $N^t = n$ får vi således likelihoodfunktionen

$$L = \prod_{i=1}^{n} \frac{\exp\left(-\int_0^{a_i} \alpha(u)\, du\right)}{M} \tag{3.3}$$

3.3.8 *Bemærkning* Denne fordeling er præcis den stationære rekurrenstidsfordeling, som er omtalt i forbindelse med fornyelsesprocesser i eksempel 1.2.11. Her afspejler fordelingen selvfølgelig den "interception bias" vi får, når vi observerer aldrene på personer i live til et kalendertidspunkt t.

3.3.9 I bemærkning B.5.10 har vi omtalt hvad vi forstår ved en ikke-parametrisk maksimum likelihood estimator. Ikke-parametrisk estimation af den betingede likelihoodfunktion 3.3 er et problem, der er behandlet i litteraturen. Problemet omhandler estimation af en aftagende tæthed, og estimatoren er kendt som Grenander estimatoren. Fra [Var89, p. 752] og [Kei91, p. 385] har jeg taget referencerne [Gre56], [Rao69], [Weg70], [BBBB72, p. 223], [BAL80], [Gro85] og [DV86], som jeg ikke selv har læst.

3.3.10 *Bemærkning* Likelihoodfunktionen (3.3) efter vi har betinget med antal personer i tværsnittet afhænger ikke af parameteren β for fødselsprocessen. Dette er klart, da vi af (3.2) ser, at N^t er sufficient for β.

3.3.11 (β ikke tidshomogen, men kendt) Hvis kun α er tidshomogen, og $\beta(t)$ afhænger af kalendertiden, bliver det ikke helt så pænt.

Likelihoodfunktionen bliver

$$L \propto \exp\left(-\int_0^\infty \beta(t-a)e^{-\int_0^a \alpha(u)\,du}\,da\right)\prod_{i=1}^{N^t}\beta(t-a_i)\exp\left(-\int_0^{a_i}\alpha(u)\,du\right)$$

og hvis vi betinger med $N^t = n$, bliver $a_1, \ldots, a_n$ iid efter fordelingen med tæthed

$$\frac{\beta(t-a)\exp\left(-\int_0^a \alpha(u)\,du\right)}{\int_0^\infty \beta(t-a)\exp\left(-\int_0^a \alpha(u)\,du\right)\,da} \tag{3.4}$$

Vi ser, at β i begge tilfælde foretager en vægtning af overlevelsessandsynlighederne. Intuitivt burde udseendet af β også i begge tilfælde få betydning. Hvis β f.eks. er nul på intervallet $[t-\Delta, t]$, kan vi ikke observere individer med alder mindre end Δ.

Hvis β er kendt, definerer vi transformationen $h(a) = \int_0^a \beta(t-s)\,ds$. Sæt $\tilde{a}_i = h(a_i)$, $i = 1, \ldots, n$. Ved transformation af fordelingen (3.4) med afbildningen h får vi, at $\tilde{a}_1, \ldots, \tilde{a}_n$ følger en fordeling med tæthed

$$\frac{\overline{F}(h^{-1}(\tilde{a}))}{\tilde{M}}$$

hvor $\tilde{M}$ er middelværdien i fordelingen med overlevelsesfunktion $\overline{F}\circ h^{-1}$. Ved samme ikke-parametriske metode som før, finder vi et estimat for $\overline{F}\circ h^{-1}$. Vi får nu umiddelbart et estimat for $\overline{F}$, da h er kendt.

En lignende metode vil formodentlig kunne anvendes i de følgende afsnit.

3.3.12 *Eksempel* (Eksponentialfordelingen) I tilfældet med tidshomogenitet og eksponentialfordelte overlevelsestider, $\beta(t) = \beta$, $\alpha(t,a) = \alpha$, bliver den ubetingede likelihoodfunktion

$$L \propto \exp\left(-\beta\frac{1}{\alpha} - \alpha\sum_{i=1}^{N^t} a_i\right)\beta^{N^t}$$

som maksimeres for $\hat{\beta} = (N^t)^2/(\sum_{i=1}^{N^t} a_i)$ og $\hat{\alpha} = N^t/(\sum_{i=1}^{N^t} a_i)$. Bemærk, at estimatoren for β kommer fra ligningen $\mathrm{E}\,N^t = \beta\frac{1}{\alpha}$ fra 2.5.16.

Den betingede likelihoodfunktion er

$$L = \prod_{i=1}^{N^t} \frac{\exp(-\alpha a_i)}{\frac{1}{\alpha}} = \prod_{i=1}^{N^t} \alpha e^{-\alpha a_i}$$

Vi ser, at de observerede aldre stammer fra den oprindelige eksponentialfordeling. Estimatet for α er det samme, som når vi ikke betinger med antal individer.

3.3.13 Ved et tværsnit i befolkningen til tid t får vi den pæneste estimation, ved at betinge med antallet af personer i live. Herved får vi iid levetider fra den stationære rekurrenstidsfordeling.

3.4 Dødsfald i kalendertidsinterval

3.4.1 I dette afsnit ser vi alle dødsfald i et kalendertidsinterval $[t_1, t_2]$. Specielt ser vi også dødsfald for personer, der fødes i intervallet.

3.4.2 I eksempel 2.4.10 så vi, at dødsfaldene i Lexis diagrammet udgør en poissonproces med intensitet

$$\lambda(t,a) = \beta(t-a)e^{-\int_0^a \alpha(t-a+u,u)\,du}\alpha(t,a) = \beta(t-a)f_{t-a}(a)$$

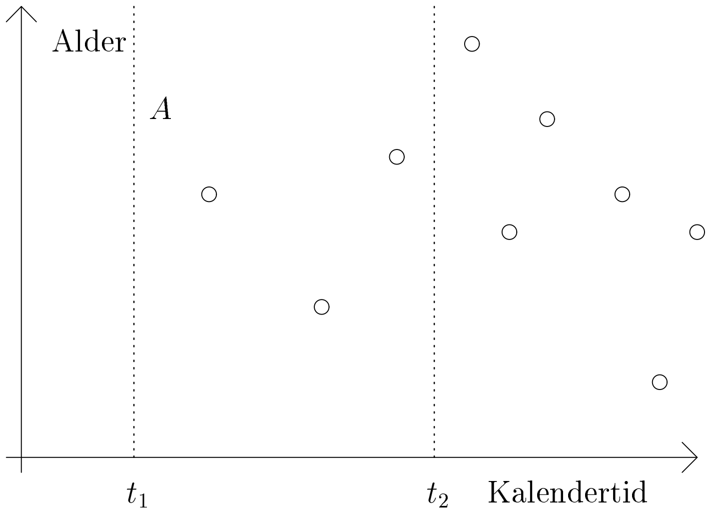
Vi observerer denne poissonproces over mængden $[t_1, t_2] \times \mathbb{R}_0$. Situationen er skitseret i figur 3.2.

3.4.3 (Parameterområde) For at opstille likelihoodfunktionen for observation af processen over $[t_1, t_2] \times \mathbb{R}_0$ skal det forventede antal dødsfald i området være endeligt. Vi vil blot bemærke, at det er tilstrækkeligt, men ikke nødvendigt, at kun endeligt mange personer er i live til tid t_1. Betingelser for dette er diskuteret i afsnit 3.2.

3.4.4 *Notation* Sæt $A = [t_1, t_2] \times \mathbb{R}_0$ og lad μ være poissonprocessen af dødsfald i Lexis diagrammet. Antag, at vi i området A observerer dødsfaldene $(t_i, a_i)_{i=1,\ldots,\mu(A)}$. Dvs. individ i dør til tid t_i i alder a_i. Vi nummererer observationerne tilfældigt som i bemærkning C.4.5. Det burde ikke give anledning til forvirring, at vi genbruger betegnelserne t_1 og t_2.

3.4.5 Hvis vi som referencemål vælger poissonprocessen med intensitet, hvor $\beta \equiv 1$, $\alpha \equiv 1$, får vi likelihoodfunktionen

$$L = \exp\left(-\int_{t_1}^{t_2}\int_0^\infty \beta(t-a)f_{t-a}(a)\,da\,dt + (t_2 - t_1)\right) \cdot \prod_{i=1}^{\mu(A)} \beta(t_i - a_i) f_{t_i - a_i}(a_i)$$



Figur 3.2: Observation af dødsfald i Lexis diagrammet i intervallet $[t_1, t_2]$. Vi ser dødsfaldene, bollerne, mellem de to lodrette stiplede linier.

I det tidshomogene tilfælde reducerer dette til

$$L \propto e^{-\beta\cdot(t_2-t_1)}\beta^{\mu(A)}\prod_{i=1}^{\mu(A)} f(a_i)$$

3.4.6 *Bemærkning* Fremover vil vi ikke skrive den fulde likelihoodfunktion op, men ignorere bidrag fra referencemålet. Referencemålet vil vi overalt tænke på som målet med samme form af intensiteten og parametrene sat til 1.

3.4.7 I det tidshomogene tilfælde skal vi estimere β ved $\hat{\beta} = \mu(A)/(t_2 - t_1)$, og α skal estimeres som om, at vi havde $\mu(A)$ observationer fra fordelingen bestemt af α. Vi kan bruge Nelson-Aalen estimatoren, som en ikke-parametrisk maksimum likelihood estimator, til at estimere den integrerede intensitet, $\int_0^{\cdot} \alpha$.

3.4.8 *Bemærkning* Vi skal estimere *som om* vi havde observationer fra $f(a)$. Det har vi imidlertid ikke umiddelbart, da individer lever på skrå linier med hældning 1, og derfor krydser det område, vi observerer over. Se også bemærkning 3.8.10 om elementærmængder for død.

3.4.9 *Bemærkning* Da vi observerer over et produktområde $A = [t_1, t_2] \times \mathbb{R}_0$ afhænger likelihoodfunktionen i det tidshomogene tilfælde ikke af kalendertidspunkterne t_i hvor dødsfaldene er fundet sted.

3.4.10 Hvis vi betinger med $\mu(A) = n$ får vi ifølge bemærkning C.4.5, at parrene (t_i, a_i) er iid observationer fra fordelingen med tæthed $\lambda(t, a)/\Lambda(A)$. Likelihoodfunktionen bliver

$$L \propto \prod_{i=1}^{n} \frac{\exp\left(-\int_0^{a_i} \alpha(u)\,du\right)\alpha(a_i)}{t_2 - t_1} = \prod_{i=1}^{n} \frac{f(a_i)}{t_2 - t_1}$$

3.4.11 *Bemærkning* Hvis vi betinger med antal observationer er parene $(t_i, a_i)_{i=1,\ldots,\mu(A)}$ uafhængige, og t_i og a_i uafhængige, t_i ligefordelt på $[t_1, t_2]$ og a_i følger levetidsfordeling med tæthed $f(a)$. Der ændres ikke på, hvordan vi skal estimere α.

3.4.12 *Bemærkning* Selv om estimatet for α ikke ændres, så vil variansen på estimatet for α afhænge af om antallet af individer er stokastisk eller fast.

3.4.13 *Eksempel* (Eksponentialfordelte levetider) Når $\beta(t) = \beta$ og $\alpha(t, a) = \alpha$ bliver $\hat{\beta} = \mu(A)/(t_2 - t_1)$ og $\hat{\alpha} = \mu(A)/(\sum_{i=1}^{\mu(A)} a_i)$. Her er $\mu(A)$ poissonfordelt $\Lambda(A) = \beta \cdot (t_2 - t_1)$ og a_i er blevet analyseret som uafhængige eksponentialfordelte med hazard α.

Estimatet for β er centralt med varians $\beta/(t_2 - t_1)$.

3.4.14 Ved $N(a) = \mu([t_1, t_2] \times [0, a])$ definerer vi en tælleproces N, der bliver en poissonproces med intensitet $\int_{t_1}^{t_2} \lambda(t, a)\,dt$. I det tidshomogene tilfælde reducerer intensiteten til $\beta \cdot (t_2 - t_1)\exp(-\int_0^a \alpha(u)\,du)\alpha(a) = \beta \cdot (t_2 - t_1) f(a)$.

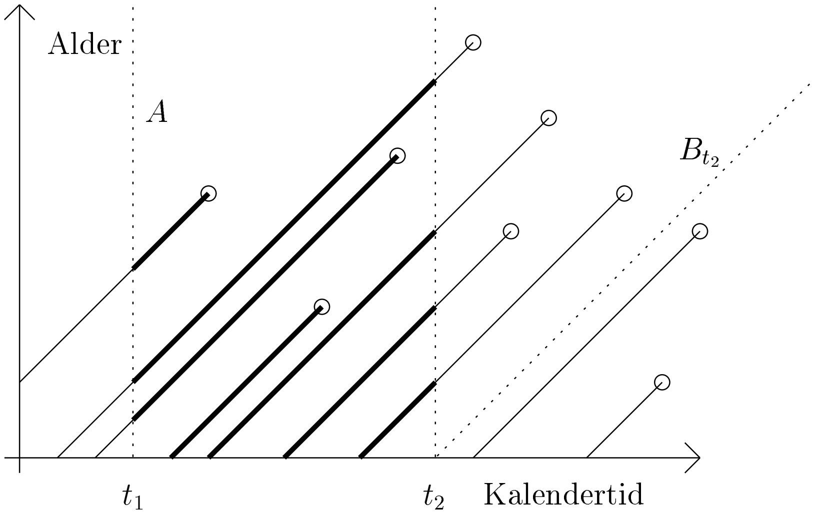
Tælleprocessen N har ikke en intensitetsproces med multiplikativ struktur, der motiverer Nelson-Aalen estimatoren som en martingalestimator.

Likelihoodfunktionen for observation af N over $\mathbb{R}_0$ bliver således mage til likelihoodfunktionen for observation af μ over mængden A. Det er værd at bemærke, at likelihoodfunktionen for observation af N kun er for observation af aldrene a_i, og ikke tidspunkterne t_i. Likelihoodfunktionen for observation af μ er for observation af parrene (t_i, a_i). Vi får her samme estimation, uanset om vi medtager kalendertidspunkterne t_i eller ej.

3.4.15 Hvis vi observerer dødsfald i et kalendertidsinterval, skal vi i det tidshomogene tilfælde analysere de observerede levetider som observationer fra fordelingen bestemt af hazarden $\alpha(a)$. Dette gælder både, hvis vi har betinget med antallet af observationer, og hvis vi ikke har. Vi kan bruge Nelson-Aalen estimatoren som en ikke-parametrisk maksimum likelihood estimator.

3.5 Personer og dødsfald i kalendertidsinterval

3.5.1 Vi observerer dødsfald og personer i live i et kalendertidsinterval $[t_1, t_2]$. Lidt løst kan vi sige, at vi ser de livslinier, der går gennem intervallet. Vi ser også personer, der fødes i intervallet. Situationen er illustreret i figur 3.3.



Figur 3.3: Observation af dødsfald og personer, der lever i kalendertidsintervallet $[t_1, t_2]$. De fede livslinier markerer hvad vi ser af personernes liv. Vi ser, hvad der sker, mellem de to lodrette stiplede linier. En bolle markerer dødsfald.

3.5.2 *Notation* Lad igen μ være poissonprocessen for dødsfald i Lexis diagrammet med intensitet

$$\lambda(t,a) = \beta(t-a)e^{-\int_0^a \alpha(t-a+u,u)\,du}\alpha(t,a) = \beta(t-a)f_{t-a}(a)$$

Definer $A = [t_1, t_2] \times \mathbb{R}_0$, og trekanten $B_{t_2} = \{(s,a) \in \mathbb{R} \times \mathbb{R}_0 \,|\, s \geqq t_2, a \geqq s - t_2\}$.

3.5.3 Vi vil tillægge det en præcis mening, at vi ser livslinierne gennem intervallet $[t_1, t_2]$. Man indser straks, at det at se en livslinie i intervallet $[t_1, t_2]$ er ækvivalent med at se øverste højre endepunktet af linien inden for A. Det af linien vi ser, og endepunktet er i entydig korrespondance. Vi kan dele vores observationer i to typer:

1. Vi observerer dødsfald i området A. Dette er gennemgået i afsnit 3.4.
2. Vi observerer en transformation af punkterne i området B_{t_2}. Vi ser nemlig præcis et tværsnit af populationen til tid t_2. Dette er gennemgået i afsnit 3.3.

De to typer af observationer er uafhængige, da de vedrører dødsfald i disjunkte mængder A og B_{t_2} af Lexis diagrammet.

3.5.4 *Notation* Lad $(t_i, a_i)_{i=1,\ldots,\mu(A)}$ være observationerne af μ i området A. Lad $\tilde{a}_1, \ldots, \tilde{a}_{\mu(B_{t_2})}$ være observationerne af aldrene i tværsnittet til tid t_2. Observationener er som før nummereret tilfældigt.

3.5.5 I 3.3.7 så vi, at vi for observationerne $\tilde{a}_1, \ldots, \tilde{a}_{\mu(B_{t_2})}$ af aldrene i tværsnittet til tid t_2 skal betinge med antallet af personer $\mu(B_{t_2})$ for at få den pæne estimation af intensiteten $\alpha(a)$.

Af symmetrigrunde vil det være naturligt, også at betinge med antallet af dødsfald $\mu(A)$ i Lexis diagrammet i mængden A. Vi så i 3.4.10, at dette ikke ændrer ved estimationen af α.

3.5.6 I det tidshomogene tilfælde skal vi bruge vores observationer til at estimere α, som uafhængige observationer — $\mu(A)$ observationer $a_1, \ldots, a_{\mu(A)}$ fra fordelingen bestemt af $\alpha(a)$, og $\mu(B_{t_2})$ observationer $\tilde{a}_1, \ldots, \tilde{a}_{\mu(B_{t_2})}$ fra den stationære rekurrenstidsfordeling hørende til α, dvs. fordelingen med tæthed $\overline{F}/M$.

3.5.7 Vi skal med andre ord maksimalisere likelihoodfunktionen

$$L \propto \prod_{i=1}^{\mu(A)} f(a_i) \prod_{i=1}^{\mu(B_{t_2})} \frac{\overline{F}(\tilde{a}_i)}{M}$$

3.5.8 Jeg har ikke fundet nogle referencer i litteraturen til dette ikke-parametriske estimationsproblem, hvor vi både har observationer fra tætheden f og den tilhørende interception biased fordeling.

I situationen med en diskret levetidsfordeling på $\mathbb{N}$ kan Vardi's RT algoritme i artiklen [Var82b] bruges. Den vil formodentlig kunne tilpasses til vores situation med kontinuerte levetidsfordelinger.

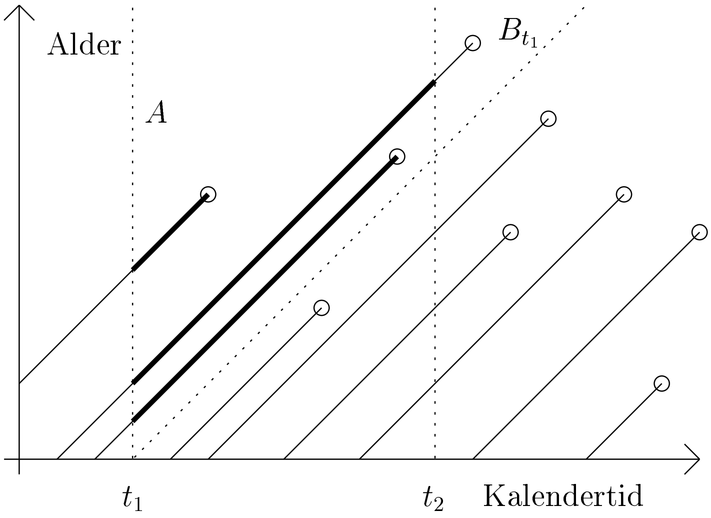
3.5.9 *Bemærkning* Vi har ovenfor betinget med $\mu(A)$ og $\mu(B_{t_2})$. Hvis vi i stedet betinger med summen $\mu(A) + \mu(B_{t_2})$ giver bemærkning C.4.5, at vi skal maksimalisere

$$L \propto \prod_{i=1}^{\mu(A)} \frac{f(a_i)}{M + (t_2 - t_1)} \prod_{i=1}^{\mu(B_{t_2})} \frac{\overline{F}(\tilde{a}_i)}{M + (t_2 - t_1)}$$

Dette er ikke så pænt som ovenfor.

3.6 Opfølgning på tværsnitsundersøgelse

3.6.1 Vi laver et tværsnit i populationen til tid t_1, og følger personerne i tværsnittet op til tid t_2. Opfølgning vil her sige, at vi til tid t_2 observerer, hvilke individer der er døde i hvilken alder og til hvilken kalendertid, samt hvilke der er i live til tid t_2. Situationen er skitseret i figur 3.4. Forskellen fra forrige afsnit er, at vi nu ikke ser individer der fødes i intervallet fra t_1 til t_2. Dette udvælgelsesmønster er det samme som bruges i [Sim80]. Vi omtaler undervejs situationen, hvor vi følger alle personer i tværsnittet til tid t_1 til deres død.



Figur 3.4: Opfølgning på tværsnitsundersøgelse. Vi følger de individer der er i live til tid t_1 op til tid t_2. Det vi ser af personernes liv er markeret med fede linier. Dødsfald er markeret med bolle.

3.6.2 *Notation* Vi benytter notation som i 3.5.2, med B_{t_1} trekanten, der har hjørne på kalendertidsaksen til tid t_1.

3.6.3 I fortsættelse af det foregående afsnit er det naturligt, at vi observerer poissonprocessen μ af dødsfald i Lexis diagrammet over mængden $A \cap B_{t_1}$, og uafhængigt heraf poissonprocessen af aldre hørende til tværsnittet til tid t_2 over mængden $[t_2 - t_1, \infty[$.

3.6.4 *Notation* Lad $(t_i, a_i)_{i=1,\dots,\mu(A\cap B_{t_1})}$ være observationerne af μ i $A \cap B_{t_1}$, og lad $\tilde{a}_1, \dots, \tilde{a}_{N^{t_2}([t_2-t_1,\infty[)}$ være observationerne i tværsnittet til tid t_2.

3.6.5 I det tidshomogene tilfælde er likelihoodfunktionen for observation af μ over $A \cap B_{t_1}$

$$L_{A\cap B_{t_1}} \propto \exp\left(-\int_{A\cap B_{t_1}} \beta f(a)\, d(t,a)\right) \prod_{i=1}^{\mu(A\cap B_{t_1})} \beta f(a_i)$$

Her er $\int_{A\cap B_{t_1}} \beta f(a)\, d(t,a) = \beta \int_0^{t_2-t_1} \overline{F}(u)\, du$. Ved betingning med antallet $\mu(A \cap B_{t_1}) = n$ får vi, at punkterne (t_i, a_i) er iid med en fordeling med tæthed

$$\frac{f(a_i)}{\int_0^{t_2-t_1} \overline{F}(u)\, du} 1\left((t_i, a_i) \in A \cap B_{t_1}\right)$$

Ifølge C.4.6 svarer dette til, i fordelingen fra 3.4.10, at betinge med, at vi er i området $A \cap B_{t_1}$. Uanset om vi betinger med antal observationer eller ej, er det ikke en pæn likelihoodfunktion.

3.6.6 I det tidshomogene tilfælde bliver likelihoodfunktionen for observation af tværsnittet til tid t_2 over mængden $[t_2 - t_1, \infty[$

$$L_{[t_2-t_1,\infty[} \propto \exp\left(-\int_{t_2-t_1}^{\infty} \beta e^{-\int_0^a \alpha(u)\,du}\, da\right) \prod_{i=1}^{N^{t_2}([t_2-t_1,\infty[)} \beta e^{-\int_0^{\tilde{a}_i} \alpha(u)\,du}$$

Hvis vi betinger med antallet $N^{t_2}([t_2 - t_1, \infty[) = n$ bliver $\tilde{a}_1, \ldots, \tilde{a}_{N^{t_2}([t_2-t_1,\infty[)}$ iid med fordeling med tæthed

$$\frac{\exp\left(-\int_0^a \alpha(u)\,du\right)}{\int_{t_2-t_1}^{\infty} \exp\left(-\int_0^a \alpha(u)\,du\right)\, da} = \frac{\overline{F}(a)}{\int_{t_2-t_1}^{\infty} \overline{F}(a)\,da}$$

for $a \geqq t_2 - t_1$. Dette er den stationære rekurrenstidsfordeling venstretrunkeret i $t_2 - t_1$, hvilket stemmer med C.4.6 og 3.3.7.

3.6.7 *Bemærkning* Hvis $t_2 - t_1$ er lille, så er $\overline{F}(u) \approx 1$ for $u \leqq t_2 - t_1$. Dette giver $\int_0^{t_2-t_1} \overline{F}(u)\,du \approx t_2 - t_1$ og $\int_{t_2-t_1}^{\infty} \overline{F}(u)\,du \approx M - (t_2 - t_1) \approx M$. Vi kan da med god tilnærmelse bruge analysen fra afsnit 3.5.

3.6.8 Den totale ubetingede likelihoodfunktion er produktet af de to ubetingede likelihoodfunktioner:

$$L \propto \exp(-\beta M)\, \beta^{\mu(B_{t_1})} \prod_{i=1}^{\mu(A\cap B_{t_1})} f(a_i) \prod_{i=1}^{N^{t_2}([t_2-t_1,\infty[)} \overline{F}(\tilde{a}_i)$$

Dette er meget pænere end det foregående, og når vi betinger med det totale antal $\mu(A \cap B_{t_1}) + N^{t_2}([t_2 - t_1, \infty[) = \mu(B_{t_1})$ får vi ifølge bemærkning C.4.5 iid observationer[1]. Den betingede likelihoodfunktion bliver

$$L_{\text{beting}} \propto \prod_{i=1}^{\mu(A\cap B_{t_1})} \frac{f(a_i)}{M} \prod_{i=1}^{N^{t_2}([t_2-t_1,\infty[)} \frac{\overline{F}(\tilde{a}_i)}{M} \tag{3.5}$$

3.6.9 *Bemærkning* Vi ser af dette, at vi skal betinge med et antal, der har en pæn forventet værdi. Det vil her sige det totale antal personer, vi har observeret, og ikke de to delantal, der tæller antal personer, som er døde hhv. i live. Dette gælder også i afsnit 3.7 og 3.8.

Jeg har ikke fundet referencer til ikke-parametrisk estimation af likelihoodfunktionen (3.5).

Likelihoodfunktionen (3.5) bliver fortolket nedenfor.

3.6.10 I det følgende gennemgås metoden i [Sim80]. Den er en svag drejning af ovenstående, og giver samme resultat. Forskellen mellem de to metoder fremhæves i 3.6.18.

[1] Med notationen her for observationerne fremgår dette ikke klart. Vi kan fra notationen, nummereringen, se om observationen stammer fra et dødsfald eller alder ved tid t_2.

3.6.11 Antag først, at vi følger alle personer i tværsnittet til tid t_1, til de dør. Vi observerer Lexis diagrammet over mængden B_{t_1}, hvor det forventede antal dødsfald i det tidshomogene tilfælde er $\int_0^\infty \int_{t_1}^{t_1+a} \beta f(a)\, dt\, da = \beta M$. βM er selvfølgelig også det forventede antal personer i live ved tværsnittet til tid t_1. Hvis tidspunkterne for dødsfald er $(t_i, a_i)_{i=1,\ldots,\mu(B_{t_1})}$ bliver likelihoodfunktionen

$$L \propto \exp(-\beta M) \prod_{i=1}^{\mu(B_{t_1})} \beta f(a_i)$$

Det er ikke umiddelbart klart, hvordan vi skal maksimere denne likelihoodfunktionen.

3.6.12 Hvis vi betinger med $\mu(B_{t_1}) = n$ bliver parrene (t_i, a_i) iid med tæthed

$$\frac{f(a_i)}{M} 1\left((t_i, a_i) \in B_{t_1}\right) \tag{3.6}$$

mht. Lebesguemålet l^2 på $\mathbb{R}^2$. I det følgende regner vi i denne fordeling.

3.6.13 Det er dette resultat [Sim80] bruger. Vores bevis for dennne fordeling er noget anderledes, da vi lader en poissonproces på hele aksen $\mathbb{R}$ beskrive fødslerne. Simon starter til tid 0 og må derfor både bruge grænseovergang for $t \to \infty$ og rækkeudviklingen af eksponentialfunktionen. Vi bruger bemærkning C.4.5 om poissonprocessen og eksempel 2.4.10.

3.6.14 *Bemærkning* Parrene (t_i, a_i) er i entydig korrespondance med sættet $(f_i, e_i) = (a_i - (t_i - t_1), t_i - t_1)$, hvor f_i er levetiden før tidspunktet t_1 for tværsnittet og e_i er levetiden efter tid t_1. Så $a_i = f_i + e_i$. Det er derfor naturligt at fortolke likelihoodfunktionen (3.6) som den simultane fordeling af en forlæns og baglæns rekurrenstid. Denne er omtalt i eksempel 1.2.11.

3.6.15 *Bemærkning* Den marginale fordeling af a_i får tæthed

$$\int 1\left((t, a) \in B_{t_1}\right) \frac{f(a)}{M}\, dt = \frac{a f(a)}{M}$$

for $a \geqq 0$. Dette er præcis den “length biased” fordeling, som er omtalt i eksempel 1.2.11.

Vi er således i en situation, hvor hverken dødeligheden eller fødselsprocessen afhænger af kalendertiden, og hvor likelihoodfunktionen kun afhænger af kalendertiden for observationen gennem indikatorfunktionen for, at vi er i det rigtige område. Alligevel bidrager observation af t_i med information, da den tillader os at splitte den “length biased” levetid a_i op i to dele. Kalendertidspunktet t_i bidrager dog ikke med mere information end, at likelihoodfunktionen stadig er den samme som funktion af parameteren α, hvis vi ser bort fra det.

3.6.16 *Bemærkning* I stil med 3.4.14 kan vi definere tælleprocessen $N(a) = \mu(B_{t_1} \cap \mathbb{R} \times [0, a])$, der tæller antal dødsfald med aldre mindre end a i området B_{t_1} af Lexis diagrammet. Vi smider information om kalendertidspunkterne t_i væk. N bliver en poissonproces med intensitet $\beta a f(a)$, hvilket nemt ses ved at udregne det forventede antal dødsfald ud i området $B_{t_1} \cap (\mathbb{R} \times [0, a])$. Hvis vi her betinger med antallet af personer i tværsnittet, ser vi igen, at de observerede aldre bliver iid med fordeling fra den "length biased" fordeling.

3.6.17 Vi observerer ikke alle dødfald i B_{t_1}, men stopper observation til tid t_2. Vi skal derfor i fordelingen (3.6) af (t_i, a_i) højrecensurere observationerne for de personer, der er i live til tid t_2.

Bemærk først, at antallet af dødsfald i området B_{t_1} er det samme som antallet af individer i live i vores tværsnit til tid t_1. Det giver således god mening, at højrecensurere observationerne. Ved højrecensur af en observation (t_i, a_i) forstår vi følgende:

Vi ser observationen, hvis $(t_i, a_i) \in A \cap B_{t_1}$. Hvis $(t_i, a_i) \in B_{t_1} \setminus A$, dvs. $t_i > t_2$, ser vi funktionen $h(t_i, a_i) = (t_2, a_i - (t_i - t_2))$, hvor andenkoordinaten er den alder personen har til kalendertid t_2. Sættes funktionen h til identiteten på mængden $B_{t_1} \cap A$ ser vi altså funktionen $h(t_i, a_i)$ af vores observationer. Dette er en fordeling med tæthed mht. Lebesguemålet l^2 på $A \cap B_{t_1}$ og mht. l på linien $\{t_2\} \times [t_2 - t_1, \infty[$ i $\mathbb{R}^2$. Vi vil finde likelihoodfunktionen for denne fordeling i en vis detalje, for senere i lignende situationer blot at henvise hertil.

Lad p være sandsynligheden for at (t_i, a_i) er i området $A \cap B_{t_1}$ og $q = 1 - p$ sandsynligheden for området $B_{t_1} \setminus A$. I den betingede fordeling af (t_i, a_i) givet vi er i området $B_{t_1} \setminus A$ finder vi nu fordelingen af $h(t_i, a_i)$, der er en simpel transformation. Fordelingen af $h(t_i, a_i) = (h_1, h_2)$ givet (t_i, a_i) er i området $B_{t_1} \setminus A$ får tæthed

$$\frac{\overline{F}(h_2)}{qM} 1\,(h_1 = t_2, h_2 > t_2 - t_1)$$

Vi kan nu sammenveje de betingede fordelinger af $h(t_i, a_i)$, givet hvilket af områderne $B_{t_1} \setminus A$ og $B_{t_1} \cap A$ vi er i, med de relevante sandsynligheder p og q. Bidraget til likelihoodfunktionen er derfor $f(a_i)/M$ hvis $(t_i, a_i) \in A \cap B_{t_1}$ og $\overline{F}(a_i - (t_i - t_2))/M$ hvis $(t_i, a_i) \in B_{t_1} \setminus A$. Den totale likelihoodfunktion bliver netop (3.5).

Bidraget ved censur svarer til at integrere tætheden (3.6) langs linien $\{(t_2 + s, h(t_i, a_i) + s) | s \in \mathbb{R}_0\}$. Bidraget ved censur er det samme som ved et tværsnit til tid t_2 og stammer fra den stationære rekurrenstidsfordeling.

3.6.18 Forskellen mellem metoden i 3.6.8 og i [Sim80] er, at

- I 3.6.8 har vi først udført transformationen i 3.3.3, og dernæst betinget med det totale antal observationer.

- I gennemgangen af [Sim80] betinger vi først med det totale antal observationer og udfører dernæst en transformation af data.

Vi har set, at vi får den samme likelihoodfunktion i de to tilfælde. Der gælder oplagt:

Lad μ være en poissonproces på $(S, \mathbb{S})$ med intensitetsmål Λ. Antag S er et borelrum og $\Lambda(S) < \infty$. Lad der være givet en afbildning $t : (S, \mathbb{S}) \to (\tilde{S}, \tilde{\mathbb{S}})$. Antag, at t opfylder betingelserne fra korollar C.4.15. Hvis vi observerer $t(\mu)$ og betinger med antallet $t(\mu)(\tilde{S}) = n$, da får vores observationer samme fordeling, som hvis vi betinger med $\mu(S) = n$ og dernæst transformerer punkterne med t.

I de følgende afsnit vil vi betinge først, og derefter transformere data.

3.6.19 *Bemærkning* Hvis vi i stedet højrecensurerede i den marginale fordeling af a_i ville vi få bidrag

$$\int_h^\infty \frac{af(a)}{M}\,da = h\frac{\overline{F}(h)}{M} + \int_h^\infty \frac{\overline{F}(u)}{M}\,du$$

hvor h er alderen ved censur. Når vi kun følger tværsnittet ved tid t_1 op til tid t_2, er det således *ikke* underordnet om vi medtager information om kalendertidspunktet t_i eller ej.

3.6.20 Når vi følger et tværsnit til tid t_1, til personerne dør eller til et kalendertidspunkt t_2, får vi en likelihoodfunktion vi kan fortolke ved at betinge med antallet af personer i live ved tid t_1. Vi skal da analysere levetiderne a_i som om, at de kom fra den simultane fordeling af (t_i, a_i) med tæthed $f(a_i)/M$. Hvis vi kun følger individerne til tid t_2, skal individer, der overlever til tid t_2, indgå i likelihoodfunktionen med en faktor $\overline{F}(h)/M$, hvor h er alderen til tid t_2, mens individer i live skal indgå med samme faktor som før.

3.7 Kohorteundersøgelse

3.7.1 I dette afsnit gennemgår vi en kohorteundersøgelse. Vi følger alle individer født i et kalendertidsinterval $[t_1, t_2]$. Situationen er skitseret i figur 3.5.

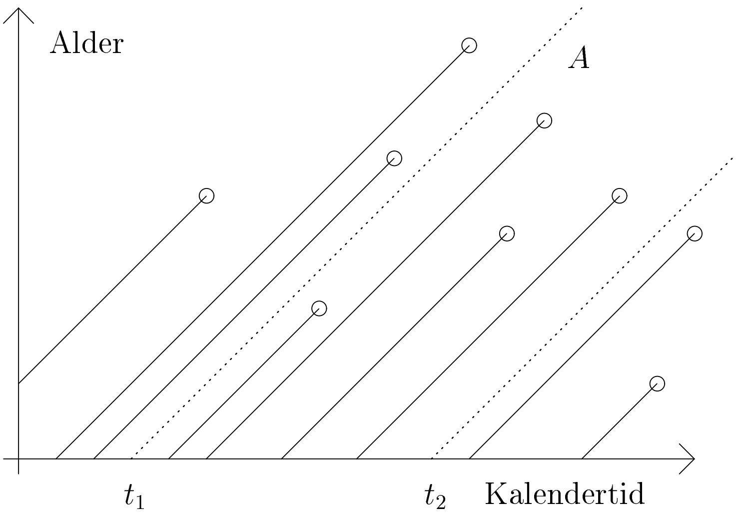
3.7.2 Vi observerer Lexis diagrammet over mængden

$$A = \{(t, a) \in \mathbb{R} \times \mathbb{R}_0 | t - a \in [t_1, t_2]\}$$

Det forventede antal dødsfald er

$$\int_A \lambda(t,a)\,d(t,a) = \int_0^\infty \int_{t_1+a}^{t_2+a} \beta(t-a) f_{t-a}(a)\,dt\,da = \int_{t_1}^{t_2} \beta(s)\,ds,$$

der selvfølgelig er det forventede antal fødsler i intervallet $[t_1, t_2]$.



Figur 3.5: Kohorteundersøgelse. Vi følger individer født i intervallet $[t_1, t_2]$. Dvs. vi observerer dødsfald i Lexis diagrammet over mængden A markeret med stiplede linier. Dødsfald er markeret med en bolle.

3.7.3 *Notation* Lad $(t_i, a_i)_{i=1,\dots,\mu(A)}$ være de observerede dødsfald i Lexis diagrammet i mængden A.

3.7.4 Likelihoodfunktionen bliver

$$L \propto \exp\left(-\int_{t_1}^{t_2} \beta(s)\, ds\right) \prod_{i=1}^{\mu(A)} \beta(t_i - a_i) f_{t_i - a_i}(a_i)$$

der i det tidshomogene tilfælde reducerer til

$$L \propto e^{-\beta\cdot(t_2-t_1)} \beta^{\mu(A)} \prod_{i=1}^{\mu(A)} f(a_i)$$

3.7.5 I det tidshomogene tilfælde skal vi estimere β ved $\hat{\beta} = \mu(A)/(t_2 - t_1)$ og α vha. $\mu(A)$ observationer fra levetidsfordelingen f. Hvis vi betinger med antallet $\mu(A) = n$, så skal vi stadig estimere α vha. observationer fra f. Dette er samme resultat som i afsnit 3.4, og vi giver en forklaring på denne sammenhæng i bemærkning 3.8.10.

3.7.6 Ved betingning med antallet af dødsfald i A, $\mu(A) = n$, bliver parrene (t_i, a_i) iid med fordeling med tæthed

$$\frac{f(a_i)}{t_2 - t_1} 1\left((t_i, a_i) \in A\right)$$

Dødsfaldene i A er netop de individer, der bliver født i $[t_1, t_2]$. Hvis vi stopper observation af kohorten til tid t_3, kan vi præcis som i 3.6.17 vise, at individer i live til tid t_3 skal højrecensureres i den alder de har til tid t_3. De censurerede individer bidrager med en faktor $\overline{F}(a_i - (t_i - t_3))$ til likelihoodfunktionen.

3.7.7 *Bemærkning* Det er interessant, at det forventede antal personer vi ser, ikke afhænger af fordelingen af levetiderne. Det betyder nemlig, at vi også skal analysere de observerede aldre som uafhængige observationer fra fordelingen f — eventuelt højrecensureret i alderen ved tid t_3 — selv om vi ikke betinger med antallet af personer $\mu(A)$. Likelihoodfunktionen bliver den samme som funktion af f, uanset om vi betinger eller ej.

3.7.8 Ved en kohorteundersøgelse skal vi analysere levetiderne fra fordelingen f, og eventuelt højrecensurere i den alder individet har ved et opfølgningstidspunkt t_3. Vi kan bruge Nelson-Aalen estimatoren.

3.8 Elementærmængder for død

3.8.1 I dette afsnit gennemgår vi, hvordan estimation for de tre typer elementærmængder for død i Lexis diagrammet skal foretages. Elementærmængderne er illustreret i figur 2.2 på side 31.

3.8.2 *Notation* Den første type elementærmængde A_1 omfatter dødsfald for individer født i intervallet $[t_1, t_2]$ og døde i aldersintervallet[2] $[a_1, a_2]$.

Lad $(t_i, a_i)_{i=1,\ldots,\mu(A_1)}$ være tidspunkter og aldre for de observerede dødsfald.

3.8.3 Det forventede antal dødsfald i denne mængde er $\int_{t_1}^{t_2} \beta(s) \int_{a_1}^{a_2} f_s(a)\, da\, ds$, der i det tidshomogene tilfælde reducerer til $\beta \cdot (t_2 - t_1)(F(a_2) - F(a_1))$. Her er $F(a_2) - F(a_1)$ blot sandsynligheden for at et individ dør i aldersintervallet $[a_1, a_2]$. Likelihoodfunktionen for observation af dødsfald i denne mængde er derfor

$$L \propto \exp\left(-\beta \cdot (t_2 - t_1)(F(a_2) - F(a_1))\right) \prod_{i=1}^{\mu(A_1)} \beta f(a_i)$$

Dette er ikke en pæn funktion af parameteren α.

3.8.4 Hvis vi betinger med antallet af dødsfald i mængden A_1, $\mu(A_1) = n$, bliver observationerne (t_i, a_i) iid med fordeling med tæthed

$$\frac{f(a_i)}{(F(a_2) - F(a_1))(t_2 - t_1)} 1\left((t_i, a_i) \in A_1\right)$$

Vi skal således analysere aldrene som venstretrunkerede i alder a_1 og højretrunkerede i alder a_2. Dette gælder også i den marginale fordeling af a_i.

[2] Det bør ikke give anledning til misforståelser, at vi bruger betegnelserne a_1, a_2, både for intervalgrænser og for aldre ved dødsfald.

3.8.5 Hvis vi også ser livslinier, der krydser A_1, men hvor individerne ikke dør i A_1, benytter vi fremgangsmåden fra afsnit 3.6 om opfølgning på en tværsnitsundersøgelse.

Det forventede antal dødsfald for individer født i intervallet $[t_1, t_2]$ og med alder $> a_1$ er $\overline{F}(a_1)(t_2 - t_1)\beta$. Dette er præcis de individer, hvis livslinie krydser A_1. Ved betingning med antallet af dødsfald er (t_i, a_i) iid med fordeling med tæthed $f(a_i)/(\overline{F}(a_1)(t_2 - t_1))$ for $(t_i, a_i) \in \{(t, a) : a > a_2, t - a \in [t_1, t_2]\}$. Vi skal således analysere disse levetider som venstretrunkerede i alder a_1. Som i 3.6.17 skal vi højrecensurere i alder a_2 for de individer, der passerer mængden A_1, men ikke dør i A_1. Vi kan her bruge Nelson-Aalen estimatoren.

3.8.6 *Notation* Den anden type elementærmængde A_2 omfatter dødsfald for individer født i intervallet $[k_1, k_2]$ og døde i kalendertidsintervallet $[t_1, t_2]$.

Lad $(t_i, a_i)_{i=1,\ldots,\mu(A_2)}$ være tidspunkter og aldre for de observerede dødsfald.

3.8.7 Det forventede antal dødsfald i A_2 er $\int_{k_1}^{k_2} \beta(s) \int_{t_1-s}^{t_2-s} f_s(a)\, da\, ds$, der i det tidshomogene tilfælde bliver $\beta \int_{k_1}^{k_2} (F(t_2 - s) - F(t_1 - s))\, ds$.

Da dette antal ikke får et pænt udseende, bliver likelihoodfunktionen ikke umiddelbart til at estimere eller fortolke. Det hjælper ikke, at betinge med antallet af individer, eller at lade $t_2 \to \infty$. Problemet er, at siderne i den kasse vi observerer over ligger "skævt" i forhold til aldersaksen.

3.8.8 *Notation* Den tredie type elementærmængde A_3 omfatter dødsfald for individer døde i kalendertidsintervallet $[t_1, t_2]$ og i aldersintervallet $[a_1, a_2]$.

Lad $(t_i, a_i)_{i=1,\ldots,\mu(A_3)}$ være tidspunkter og aldre for de observerede dødsfald.

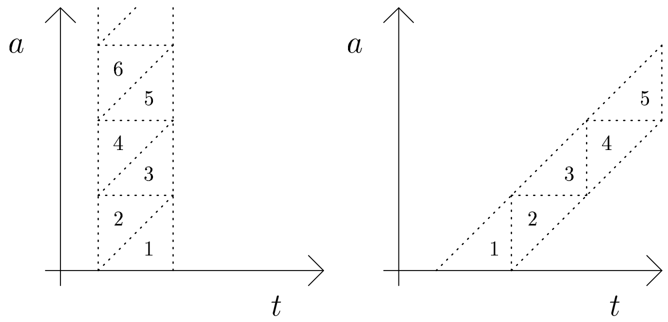
3.8.9 Der forventede antal dødsfald i A_3 er $\int_{t_1}^{t_2} \int_{a_1}^{a_2} \beta(t - a) f_{t-a}(a)\, da\, dt$, der i det tidshomogene tilfælde bliver $\beta \cdot (t_2 - t_1)(F(a_2) - F(a_1))$.

Hvis vi betinger med antallet af dødsfald vi ser i A_3, $\mu(A_3) = n$, så bliver parrene (t_i, a_i) iid med fordeling med tæthed

$$\frac{f(a_i)}{(F(a_2) - F(a_1))(t_2 - t_1)} 1\left((t_i, a_i) \in A_3\right)$$

Vi skal analysere levetiderne som både venstre- og højretrunkerede.

3.8.10 *Bemærkning* De følgende bemærkninger er kun relevante i det tidshomogene tilfælde. Når vi observerer dødsfald i elementærmængder af type 1 og 3, skal vi analysere levetiderne ens. Det er fordi mængderne er opbygget af elementærtrekanter, der svarer til hinanden. To elementærtrekanter svarer til hinanden, hvis de er ens i størrelse, type og placering i aldersretningen. Vi skal kunne komme fra den ene til den anden ved en forskydning i kalendertidsretningen. To elementærtrekanter, der svarer til hinanden, har det samme forventede antal dødsfald, og bidrager til likelihoodfunktionen på samme måde. Af samme grund skal vi analysere dødsfald i et kalendertidsinterval på samme måde som i en kohorteundersøgelse. Situationen er skitseret i figur 3.6. I diskussionen [Chi86] af Brillingers artikel [Bri86] er der nogle lignende overvejelser.



Figur 3.6: De to mængder opfører sig ens, når vi skal estimere i det tidshomogene tilfælde og kun ser dødsfald. Trekanter med samme nummer giver samme bidrag til likelihoodfunktionen. Tilsvarende opfører elementærmængder af type 1 og 3 sig ens, da de er opbygget af de samme elementærtrekanter.

3.8.11 Vi kan ikke pege på en pæn analyse af elementærmængder af type 3 i det tilfælde, hvor vi også ser livslinier, der krydser mængden. Det skyldes, at det forventede antal livslinier, der krydser mængden, ikke bliver pænt. Mere præcist bliver det forventede antal af livslinier, der passerer gennem hele mængden ikke pænt, da den "øvre kant" er bukket med form $\neg$. (Det forventede antal dødsfald i mængden blev pænt.)

3.8.12 Vi får en pæn analyse med venstretrunkering og højrecensur for elementærmængder af den første type, hvis vi både ser dødsfald og individer i live. Ser vi kun dødsfald, får vi både venstre- og højretrunkering.

For elementærmængder af den anden type kan vi ikke udtale os om en pæn estimation.

Den tredie type elementærmængde opfører sig som den første type, når vi kun ser dødsfald. Hvis vi også ser individer i live i mængden, får vi ikke en pæn analyse.

3.9 Generelt om at betinge med de fødselstidspunkter vi ser

3.9.1 I dette afsnit betinger vi med fødselstidspunkterne for de dødsfald/livslinier vi ser, så det eneste stokastiske er længden af livet. Analysen afhænger da ikke længere af antagelsen om, at fødslerne følger en poissonproces, og vi er ikke, som i det forgående, afhængige af, at det forventede antal bliver "pænt".

3.9.2 Vi bruger den generelle ramme fra afsnit 2.4, hvor fødselsprocessen η er på et rum $(S, \mathbb{S})$ med intensitetsmål Φ, og P er en familie af markovkerner fra S til et rum af livsforløb $(E, \mathcal{E})$. Vi betragter punktprocessen $\mu = \sum_{i \in I} \varepsilon_{(\sigma_i, X_i)}$ på $S \times E$.

3.9.3 *Notation* Lad $(h_s)_{s \in S}$ være en familie af målelige funktioner $h_s : E \to \{0, 1\}$.

3.9.4 Fortolkningen af $(h_s)_{s \in S}$ er følgende: for et individ født til tidspunkt s angiver h_s om vi ser individet eller ej. For $x \in E$ er

$$h_s(x) = \begin{cases} 1 & \text{hvis vi ser noget information om } x \\ 0 & \text{hvis vi ikke ser noget information om } x \end{cases}$$

3.9.5 **Sætning** *Givet σ_i, for de $i \in I$ hvor $h_{\sigma_i}(X_i) = 1$, er $(X_i)_{i : h_{\sigma_i}(X_i)=1}$ uafhængige, og X_i følger fordeling $P_{\sigma_i}(\,\cdot\,|h_{\sigma_i}(X_i) = 1)$.*

Bevis Givet $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$ er $(X_i)_{i \in I}$ uafhængige og $X_i \sim P_{\sigma_i}(\,\cdot\,)$.

Beting nu både med η og yderligere familien af funktioner $(h_{\sigma_i}(X_i))_{i \in I}$. Så er $(X_i)_{i \in I}$ uafhængige og $X_i \sim P_{\sigma_i}(\,\cdot\,|h_{\sigma_i}(X_i))$. Dette ses ved at betragte en vilkårlig endelig delmængde $\tilde{I}$ af I og benytte princippet om yderligere betingning[3]. Lad $\otimes$ angive uafhængighed. Da

$$\mathcal{L}\Bigg((X_i)_{i \in \tilde{I}} \Bigg| \eta = \sum_{i \in I} \varepsilon_{\sigma_i}\Bigg) = \otimes_{i \in \tilde{I}} P_{\sigma_i}(X_i \in \,\cdot\,)$$

får vi

$$\mathcal{L}\Bigg((X_i)_{i \in \tilde{I}} \Bigg| \eta = \sum_{i \in I} \varepsilon_{\sigma_i}, (h_{\sigma_i}(X_i))_{i \in I}\Bigg) = \otimes_{i \in \tilde{I}} P_{\sigma_i}(X_i \in \,\cdot\,|h_{\sigma_i}(X_i))$$

hvor uafhængigheden bevares, da den yderligere betingning har en produktstruktur, hvor der sættes krav til hver koordinat X_i for sig.

Givet η og $(h_{\sigma_i}(X_i))_{i \in I}$ er $(X_i)_{i : h_{\sigma_i}(X_i)=1}$ uafhængige og X_i følger fordeling $P_{\sigma_i}(\cdot\,|h_{\sigma_i}(X_i) = 1)$, da vi blot tager marginal fordeling ovenfor.

Den betingede fordeling afhænger ikke af σ_i og $h_{\sigma_i}(X_i)$ for $i \in I$, hvor $h_{\sigma_i}(X_i) = 0$. Det betyder, at vi kan udelade disse fra betingningen uden at fordelingen ændres. Det ønskede følger. $\square$

3.9.6 *Bemærkning* Sætningen gælder selvfølgelig generelt for funktioner $h : S \times E \to (F, \mathcal{F})$, hvor betingelsen er, at $h(\sigma_i, x_i) \in A$ for en mængde $A \in \mathcal{F}$.

[3] Princippet om yderligere betingning siger lidt løst følgende. Lad η, X og $h(\eta, X) = Y$ være stokastiske variable. Indfør notationen $P(\cdot|\eta = z) = P_z(\cdot)$. Så er $P(\cdot|\eta = z, Y = y) = P_z(\cdot|Y = y)$, som her er $P_z(\cdot|h(z, X) = y)$. En reference er [Jac82, app. 1].

3.9.7 *Bemærkning* Sætningen gælder, uanset om fødselsprocessen η er en poissonproces eller ej. Det vigtige er, at givet fødselsprocessen er individernes livsforløb uafhængige, og deres fordeling afhænger kun af fødselstidspunktet.

Dette er en fordel, hvis vi er i tvivl om, hvorvidt fødselsprocessen er en poissonproces. Vi bygger da ikke vores model på denne antagelse.

Hvis fødslerne følger en poissonproces, vil vi forvente, at vi får en svagere analyse. Vi udnytter ikke al information i vores model, da betingningen gør noget af informationen ikke stokastisk. Lidt løst kan vi sige: "jo mere model for data, jo bedre".

3.9.8 I de næste afsnit vil vi anvende sætningen i forskellige situationer. Fælles for disse afsnit er, at vi ofte vil se forskellige former for trunkering og censur af levetiderne. I alle tilfælde er trunkerings- og censuraldrene deterministiske, når fødselstidspunktet er kendt. Der vil derfor være tale om uafhængig venstretrunkering og højrecensur. Se [ABGK93, afsn. III.2, III.3] og afsnit B.5.

3.10 Beting med alle fødselstidspunkter — tværsnit

3.10.1 Vi er i den lidt hypotetiske situation, at vi har set alle fødselstidspunkter før tid t_0 og betinger med disse. Spørgsmålet er, hvad vi kan sige vha. et tværsnit til tid t_0.

3.10.2 *Notation* Lad $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$ være fødselsprocessen og X_i levetiden hørende til fødselstidspunkt σ_i. X_i følger fordelingen med overlevelsesfunktion $\overline{F}_{\sigma_i}$.

Ifølge bemærkning C.2.7 er I højst numerabel. Antag $I = \mathbb{Z}$. Nummerer fødselstidspunkterne, så $\cdots < \sigma_i < \sigma_{i+1} < \cdots$ og σ_0 er det første fødselstidspunkt efter tid t_0.

3.10.3 Med $h_s(x) = 1\,(s \leqq t_0)$ giver sætning 3.9.5, at levetiderne X_i, $i < 0$, er uafhængige med fordeling $X_i \sim F_{\sigma_i}$, givet σ_i, $i < 0$.

3.10.4 Givet fødselstidspunkterne er det eneste vi ser til tidspunkt t_0, hvorvidt individet født til tid σ_i, $i < 0$, har overlevet op til tid t_0. Dvs. om $X_i < t_0 - \sigma_i$ (død ved tid t_0) eller $X_i \geqq t_0 - \sigma_i$ (lever ved tid t_0). Vi har således en form for censur, hvor alle levetider enten højre- eller venstrecensureres.

3.10.5 Likelihoodfunktionen for observation af individ i er derfor

$$\overline{F}_{\sigma_i}(t_0 - \sigma_i)^{1(X_i > t_0 - \sigma_i)}(1 - \overline{F}_{\sigma_i}(t_0 - \sigma_i))^{1(X_i \leqq t_0 - \sigma_i)}$$

Ifølge [Jac95b], kapitlet om følger af likelihoodfunktioner, har vi en likelihoodfunktion for observationen til tid t_0, hvis produktet

$$\prod_{i:\sigma_i \in [s,t_0]} \overline{F}_{\sigma_i}(t_0 - \sigma_i)^{1(X_i > t_0 - \sigma_i)}(1 - \overline{F}_{\sigma_i}(t_0 - \sigma_i))^{1(X_i \leqq t_0 - \sigma_i)} \tag{3.7}$$

af uafhængige faktorer konvergerer mod et positivt tal med positiv sandsynlighed for $s \to -\infty$.

3.10.6 Vi vil diskutere konvergens af (3.7) i det tidshomogene tilfælde.

3.10.7 **Sætning** (Tilstrækkelige betingelser) *Følgende betingelser er hver for sig tilstrækkelige for at produktet* (3.7) *er konvergent med positiv grænse for* $s \to -\infty$.

1. *Levetiderne er opadtil begrænset af* C, $\overline{F}(x) = 0$ *for* $x > C$.
2. *Summen* $\sum_{i=-1}^{-\infty} \overline{F}(t_0 - \sigma_i)$ *er konvergent.*

Bevis

1. Fra et vist trin i grænseovergangen får vi ns. kun 1-taller i produktet.
2. Antallet af individer i live til tid t_0 er $\sum_{i=-1}^{-\infty} 1\,(X_i > t_0 - \sigma_i)$, som har endelig middelværdi ifølge antagelsen. Dvs. der ns. kun er endeligt mange individer i live til tid t_0.

   Fra et vist trin i grænseovergangen $s \to -\infty$ kommer der således kun faktorer i produktet af formen $1 - \overline{F}(t_0 - \sigma_i)$, og så er produktet netop konvergent[4] med grænse > 0, hvis $\sum_{i=-1}^{-\infty} \overline{F}(t_0 - \sigma_i) < \infty$. □

3.10.8 *Bemærkning* Da fordelingen $\overline{F}$ har middelværdi er den anden betingelse et krav til, at der ikke fødes for mange personer, når $s \to -\infty$. Det vigtige er, at der ns. kun er endeligt mange individer i live til tid t_0. Når fødslerne følger en poissonproces har vi diskuteret betingelser for dette i afsnit 3.2.

3.10.9 I [Kei91, p. 386] er der referencer til hvordan, vi skal estimere under denne form for census, hvor vi kun ser, om et individ bliver yngre eller ældre end en given alder. Jeg har ikke læst disse referencer, der er [ABE+55] og [Gro87].

3.11 Beting med de fødselstidspunkter vi ser — tværsnit

3.11.1 Vi vil i dette afsnit behandle den situation, at vi ser et tværsnit til tid t_0 og betinger med fødselstidspunkterne for de individer vi ser. Det viser sig at være ret uinteressant.

3.11.2 Vi ser til tid t_0 individer i live i aldre a_i, $i = 1, \ldots, N^{t_0}$. Betinger vi med de tilhørende fødselstidspunkter $\sigma_i = t_0 - a_i$, så giver sætning 3.9.5 med $h_s(x) = 1\,(s \leqq t_0, x \geqq t_0 - s)$, at de tilhørende levetider X_i er uafhængige, og X_i skal analyseres i den betingede fordeling givet $X_i \geqq a_i$. Desværre er dette også alt, vi ser om X_i, så likelihoodfunktionen bliver $L = 1$.

[4] For $p_i \in]0,1[$ med $i \in \mathbb{N}$ er $\prod_{i=1}^{\infty}(1 - p_i) \in]0,1[\Leftrightarrow \sum_{i=1}^{\infty} p_i < \infty$

3.11.3 Vi er altså i en situation, hvor vi ikke kan sige noget ud fra tværsnittet, da vi har betinget med det vi ser. For at kunne sige mest muligt ud fra et tværsnit, skal vi bruge en model for fødslerne som i afsnit 3.3.

3.12 Beting med fødselstidspunkter og Nelson-Aalen estimatoren

3.12.1 I dette kapitel viser vi, at vi kan analysere levetiderne med venstretrunkering og højrecensur, når vi betinger med fødselstidspunkterne for de individer vi ser. Dette passer godt ind i rammen af tælleprocesser, og vi kan bruge Nelson-Aalen estimatoren. Hvis vi kun ser dødsfald, får vi både venstre- og højretrunkering.

3.12.2 Lad $A \subseteq \mathbb{R} \times \mathbb{R}_0$ være en målelig mængde, der opfylder:

- For alle $p_1, p_2 \in A$, hvor der findes $u > 0$ så $p_2 = p_1 + (u, u)$, skal gælde

$$\forall v \in [0,1] : vp_1 + (1-v)p_2 \in A$$

  Dette er en konveksitetsbetingelse på linier med hældning 1.

- Det forventede antal livslinier, der rammer mængden A er endeligt.

3.12.3 **Definition** $a_{1,s} = \inf_{u>0}\{(s+u,u) \in A\}$ og $a_{2,s} = \sup_{u>0}\{(s+u,u) \in A\}$.

3.12.4 $a_{1,s}$ er den alder et individ født til tid s har, når livslinien rammer mængden A, og $a_{2,s}$ er den alder individet har, når livslinien forlader mængden. Hvis der tages inf og sup over tomme mængder er $a_{1,s} = \infty$ og $a_{2,s}$ lades udefineret.

3.12.5 *Bemærkning* $a_{1,s}$ og $a_{2,s}$ er deterministiske. Vi har derfor uafhængig venstretrunkering og højrecensur i det følgende.

3.12.6 *Notation* Lad (t_i, a_i), $i = 1, \ldots, \mu(A)$, være dødsfaldene i mængden A.

3.12.7 Hvis vi betinger med fødselstidspunkterne $\sigma_i = t_i - a_i$ for individer, der dør i mængden A, så giver sætning 3.9.5 med $h_s(x) = 1\left((s+x,x) \in A\right)$, at vi skal analysere levetiderne som uafhængige. Levetiden a_i skal analyseres i fordelingen med tæthed f_{σ_i}, hvor vi venstretrunkerer i alder a_{1,σ_i} og højretrunkerer i alder a_{2,σ_i}. Bidraget til likelihoodfunktionen for individ i bliver

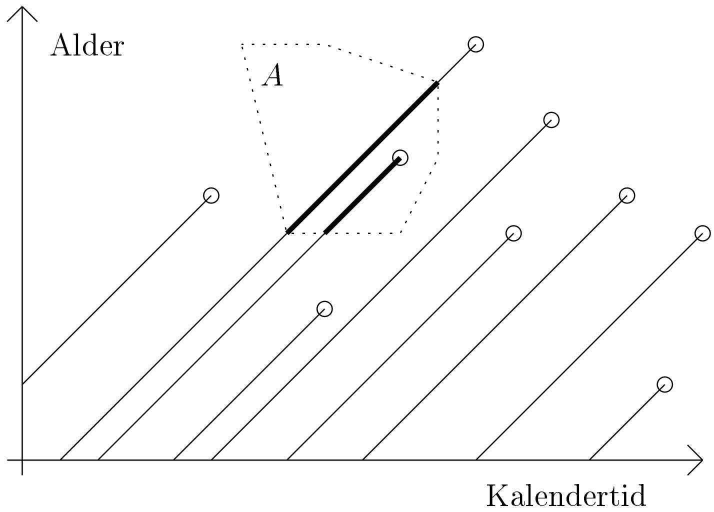
$$\frac{f_{\sigma_i}(a_i)}{F_{\sigma_i}(a_{2,\sigma_i}) - F_{\sigma_i}(a_{1,\sigma_i})} 1\left(a_i \in [a_{1,\sigma_i}, a_{2,\sigma_i}]\right)$$

3.12.8 Når vi betinger med fødselstidspunkterne $\sigma_i = t_i - a_i$ for individer, hvis livslinie krydser mængden A, siger sætning 3.9.5 med $h_s(x) = 1\left(x > a_{1,s}\right)$, at vi skal analysere de tilhørende levetider som uafhængige og venstretrunkerede i alder a_{1,σ_i} i fordelingen f_{σ_i}.

3.12.9 Når vi kun ser individer over mængden A, skal vi højrecensurere individer i live ved udgang fra mængden. Censureringsalderen er a_{2,σ_i} for individer født til tid σ_i.

3.12.10 Situationen er skitseret i figur 3.7. Vi følger individer på kalendertidsaksen (vandret), men er interesseret i at analysere levetider på aldersaksen (lodret). Dette er meget lig situationen i kapitel 1, hvor vi byttede rundt på de to tidsakser i en fornyelsesproces. Vi regner her direkte på aldersaksen, uden at definere en formel tidstransformation.

Eksempel B.5.7 giver direkte, at vi i tilfældet, hvor levetidsfordelingen ikke afhænger af kalendertiden, kan definere passende tælleprocesser med en multiplikativ struktur af intensiteten. Vi kan derfor benytte Nelson-Aalen estimatoren, og har også motiveret den vha. martingalargumenter.



Figur 3.7: Vi observerer individer i live i mængden A markeret med stiplede linier, og ser hvad der er markeret med fede linier. Dødsfald er markeret med bolle. Vi skal venstretrunkere i alder ved indgang i mængden A og højrecensurere i alder ved udgang fra mængden A, når vi betinger med fødselstidspunkterne.

3.12.11 De gennemgåede måder at sample på fra afsnit 3.4–3.8 er alle dækket af denne ramme:

Kalendertidsinterval Vi skal bruge $A = [t_1, t_2] \times \mathbb{R}_0$.

Opfølgning på tværsnit Hvis tværsnittet laves til tid t_0 skal vi bruge mængden $A = \{(t, a) \in \mathbb{R} \times \mathbb{R}_0 | t \geqq t_0, a \geqq t - t_0\}$.

Kohorteundersøgelse Vi skal bruge mængden $A = \{(t, a) \in \mathbb{R} \times \mathbb{R}_0 | t - a \in [t_1, t_2]\}$, når vi betragter kohorten født i intervallet $[t_1, t_2]$.

Elementærmængder for død anvendes umiddelbart.

3.13 Konklusion

3.13.1 Når fødselstidspunkterne følger en homogen poissonproces, og fordelingen af levetiderne ikke afhænger af kalendertiden, har vi undersøgt i hvilke fordelinger vi skal estimere levetiden, når vi sampler individer i Lexis diagrammet. Vi har set, at vi skal analysere de observerede aldre som uafhængige. I de fleste tilfælde skal vi betinge med det totale antal individer vi ser, for at få en pæn fordeling at estimere i. Lad M være middellevetiden i levetidsfordelingen. Vi skal analysere levetiderne fra en fordeling, der afhænger af måden vi sampler på:

- Ved et tværsnit skal vi analysere aldrene fra den stationære rekurrenstidsfordeling med tæthed $\overline{F}(a)/M$.
- Ved observation over et kalendertidsinterval skal dødsfald bidrage med $f(a)$, og hvis vi også ser individer, der lever gennem intervallet, skal de bidrage med $\overline{F}(a)/M$ fra alderen, hvor de forlader intervallet.
- Ved opfølgning på et tværsnit skal dødsfald bidrage med $f(a)/M$ til likelihoodfunktionen. Individer i live ved slutningen af opfølgningsperioden højrecensureres og bidrager med $\overline{F}(a)/M$.
- I en kohorteundersøgelse skal aldrene analyseres fra fordelingen $f(a)$ — eventuelt højrecensureret i alder ved opfølgningstidspunktet.
- Ved en gennemgang af elementærmængderne for død, fandt vi, at det centrale er hvilke elementærtrekanter mængden er opbygget af. Når vi kun betragter dødsfald, gør det ikke noget, hvis de er skubbet i kalendertidsretningen.

3.13.2 Hvis fødselsprocessen er inhomogen, men intensiteten er kendt, har vi i 3.3.11 set, hvordan vi kan modificere ovenstående.

3.13.3 I modellen hvor fødslerne følger en poissonproces er det nødvendigt, at det forventede antal dødsfald eller livslinier i den mængde, vi observerer, over er pænt.

3.13.4 Når vi betinger med fødselstidspunkterne for de personer vi ser, afhænger analysen ikke af fordelingen af den underliggende fødselsproces. Endvidere skal vi bruge venstretrunkering og højrecensur, når vi ser alle individer i live i en delmængde af Lexis diagrammet. Vi kan da bruge tælleprocesteknikker og Nelson-Aalen estimatoren. Hvis vi kun ser dødsfald, får vi både venstre- og højretrunkering.

3.13.5 Vi har interesseret os for det ikke-parametriske tilfælde. Det parametriske tilfælde ligger lige for, da vi har en likelihoodfunktion for observationerne. Vi har ikke set på hvilke problemer, dette kan give i konkrete eksempler.

Appendix A

Notation

Nedenfor er gengivet notation og symboler, som er brugt.

iid Uafhængige identisk fordelte.

$\mathbb{R}_+$ $]0,\infty[$

$\mathbb{R}_0$ $[0,\infty[$

$\overline{\mathbb{R}}_+$ $]0,\infty]$

$\overline{\mathbb{R}}_0$ $[0,\infty]$

$\mathbb{B}$ Borelalgebraen på $\mathbb{R}$. Borelalgebraen på de foregående delmængder af $\mathbb{R}$ betegnes med $\mathbb{B}_+$, $\mathbb{B}_0$, $\overline{\mathbb{B}}_+$ og $\overline{\mathbb{B}}_0$.

l Lebesguemålet på $\mathbb{R}$

l^n Lebesguemålet på $\mathbb{R}^n$

$\mathbb{N}_0$ $\mathbb{N} \cup \{0\}$

$\overline{\mathbb{N}}_0$ $\mathbb{N}_0 \cup \{\infty\}$

$n^{(k)}$ n i k rund, $n^{(k)} = \frac{n!}{(n-k)!}$, for $0 \leqq k \leqq n$.

$1(\cdot)$ Indikatorfunktionen for, at betingelsen $\cdot$ er opfyldt.

ε_x Etpunktsmålet i punktet x. Rummet x ligger i vil være givet fra sammenhængen.

$y.$ $y. = \sum_{i \in I} y_i$

$U_{(m)}$ $U_{(1)}, \ldots, U_{(n)}$ er de ordnede observationer af $U_1, \ldots, U_n$. I afsnit 1.7 er $V_{(1)}, \ldots, V_{(n+1)}$ de ordnede observationer af $V_1, \ldots, V_n, V_{n+1}^c$.

μ Stokastisk mål. Generel punktproces eller mærket punktproces. I kapitel 2 processen $\mu = \sum_{i \in I} \varepsilon_{(\sigma_i, X_i)}$ og i kapitel 3 processen af dødsfald i Lexis diagrammet.

N Tælleproces.

λ_s^y Intensitet for mærket punktproces for mærket y til tid s.

$P^{(n)}$ Markovkernerne $P^{(n)}_{t_1 \ldots t_n y_1 \ldots y_n}$ bestemmer næste tidspunkt i en mærket punktproces.

$\pi^{(n)}$ Submarkovkernerne $\pi^{(n)}_{t_1 \ldots t_n t\, y_1 \ldots y_n}$ bestemmer mærket hørende til tid t.

$\Lambda(\cdot)$ Intensitetsmål for poissonproces

$\lambda(\cdot)$ Intensitet for poissonproces. Dvs. tæthed for Λ

η $\eta = \sum_{i \in I} \varepsilon_{\sigma_i}$, processen af fødselstidspunkter.

Φ Intensitetsmål for η. Oftest er $\Phi = \beta \cdot l$.

$\beta(t)$ Intensitet i poissonproces, der beskriver fødsler

$\alpha(t, a)$ Dødsintensitet i Lexis diagrammet. Individet født til tid s har hazard $\alpha(s + a, a)$, $a \geqq 0$.

$P(s, \cdot)$ Markovkerne fra $(S, \mathbb{S})$ til $(E, \mathcal{E})$. Skrives også $P_s(\cdot)$.

ns Næsten sikkert, med sandsynlighed 1.

na Næsten alle. Der vil altid være angivet et mål, f.eks. κ, så κ-målet af mængden hvor udsagnet ikke gælder er nul.

$\int_0^{\cdot} f$ Hvis f er en funktion fra $\mathbb{R}_0$ skriver vi kort $\int_0^{\cdot} f$ for funktionen $t \mapsto \int_0^t f(x)\, dx$.

Appendix B

Mærkede punktprocesser

B.1 Introduktion

B.1.1 Dette appendix gengiver de resultater fra teorien om mærkede punktprocesser, som vi skal bruge. Endvidere indføres den notation, der benyttes i kapitel 1. Resultater og notation er taget fra [Jac95a].

B.2 Definition af mærkede punktprocesser

B.2.1 **Definition** En *simpel punktproces* er en følge $T = (T_n)_{n \geqq 1}$ af stokastiske variable med værdier i $\overline{\mathbb{R}}_+$, der opfylder

$$P(0 < T_1 \leqq T_2 \leqq \dots) = 1 \tag{B.1}$$

$$P(\lim_{n\to\infty} T_n = \infty) = 1 \tag{B.2}$$

$$P(T_n < T_{n+1}, T_n < \infty) = P(T_n < \infty) \tag{B.3}$$

B.2.2 Lad $(E, \mathcal{E})$ være et måleligt rum; *mærkerummet*. Definer det udvidere mærkerum $\overline{E}$ ved $\overline{E} = E \cup \{\Delta\}$, hvor $\Delta \notin E$ er det *irrelevante mærke*.

B.2.3 **Definition** En *mærket punktproces* er en følge $(T, Y) = (T_n, Y_n)_{n \geqq 1}$ af stokastiske variable med værdier i $\overline{\mathbb{R}}_+ \times \overline{E}$, hvor T er en simpel punktproces, og der yderligere gælder

$$P(Y_n \in E, T_n < \infty) = P(T_n < \infty) \tag{B.4}$$

$$P(Y_n = \Delta, T_n = \infty) = P(T_n = \infty) \tag{B.5}$$

B.2.4 **Definition** (Tælleprocessti-rummet) Rummet W af tælleprocesstier er rummet af funktioner $w : \mathbb{R}_0 \to \mathbb{N}_0$, der er cadlag og opfylder $w(0) = 0$, samt $\Delta w(t) \in \{0, 1\}$ for alle $t \in \mathbb{R}_0$.

B.2.5 **Definition** På tælleprocessti-rummet W induceres σ-algebraen $\mathcal{H}^\circ = \sigma(N_t^\circ)_{t \geqq 0}$ af projektionerne, og den tilhørende filtrering $\mathcal{H}_t^\circ = \sigma(N_s^\circ)_{0 \leqq s \leqq t}$ for $t \in \mathbb{R}_0$.

B.2.6 **Definition** En *tælleproces* er en stokastisk variabel N med værdier i $(W, \mathcal{H}^\circ)$.

B.2.7 Hvis T er en simpel punktproces, er den tilhørende tælleproces N givet ved $N_t = N_t^\circ \circ N$, hvor $N_t = \sum_{n=1}^\infty 1\,(T_n \leqq t)$.

B.2.8 **Definition** Filtreringen $(\mathcal{F}_t^N)_{t \geqq 0}$ på Ω induceret af N er givet ved $\mathcal{F}_t^N = \sigma(N_s)_{0 \leqq s \leqq t}$.

B.2.9 *Notation* Sæt $S = \mathbb{R}_0 \times E$ og $\mathbb{S} = \mathbb{B}_0 \otimes \mathcal{E}$.

B.2.10 **Definition** Mængden M^d af *diskrete tællemål* på S er mængden af mål m på $(S, \mathbb{S})$, der opfylder

- $m(A) \in \overline{\mathbb{N}}_0$ for alle $A \in \mathbb{S}$.
- Der findes $C \in \mathbb{S}$ som er højst tællelig, så $m(\complement C) = 0$.

Lad M betegne mængden af mål $m \in M^d$, der opfylder

- $m(\{0\} \times E) = 0$
- $m([0, t] \times E) < \infty$ for alle $t \in \mathbb{R}_0$
- $m(\{t\} \times E) \in \{0, 1\}$ for alle $t \in \mathbb{R}_0$

B.2.11 **Definition** σ-algebraen $\sigma(M)$ på M induceres initialt af familien af afbildninger $(m \mapsto m(A))_{A \in \mathbb{S}}$ fra M til $\overline{\mathbb{N}}_0$.

B.2.12 **Definition** Et *tilfældigt tællemål* er en stokastisk variabel μ med værdier i $(M, \sigma(M))$.

B.2.13 **Sætning** *En MPP (T, Y) er i entydig korrespondance med det tilfældige tællemål*

$$\mu = \sum_{\substack{n \in \mathbb{N} \\ T_n < \infty}} \varepsilon_{(T_n, Y_n)}$$

B.2.14 Ud fra en MPP (T, Y) definerer vi tælleprocessen $N(A)$ ved

$$N_t(A) = \mu([0, t] \times A) = \sum_{n=1}^\infty 1\,(T_n \leqq t, Y_n \in A)$$

$N(A)$ tæller antal mærker i A.

B.2.15 **Definition** Filtreringen $(\mathcal{F}_t^\mu)_{t \geqq 0}$ induceret af en mærket punktproces (T, Y) (eller μ) er defineret ved $\mathcal{F}_t^\mu = \sigma(N_s(A))_{0 \leqq s \leqq t, A \in \mathcal{E}}$.

B.3 Fordelingen af en mærket punktproces

B.3.1 **Definition** Fordelingen af en simpel punktproces T er den transformerede sandsynlighed $Q^\circ = N(P)$ på $(W, \mathcal{H}^\circ)$.

B.3.2 **Definition** Fordelingen af en mærket punktproces (T, Y) er den transformerede sandsynlighed $Q^\circ = \mu(P)$ på $(M, \sigma(M))$.

B.3.3 **Sætning** *Fordelingen P° af den mærkede punktproces (T, Y) er bestemt ved fordelingen $P^{(0)}$ af T_1° og de regulære betingede fordelinger $P^{(n)}_{T_1^\circ,\ldots,T_n^\circ,Y_1^\circ,\ldots,Y_n^\circ}(\cdot)$ af T_{n+1}° givet $(T_1^\circ, \ldots, T_n^\circ, Y_1^\circ, \ldots, Y_n^\circ)$, samt de regulære betingede fordelinger $\pi^{(n)}_{T_1^\circ,\ldots,T_{n+1}^\circ,Y_1^\circ,\ldots,Y_n^\circ}(\cdot)$ af Y_{n+1}° givet $(T_1^\circ, \ldots, T_{n+1}^\circ, Y_1^\circ, \ldots, Y_n^\circ)$.*

B.3.4 *Notation* Vi bruger notation som i sætningen.

B.4 Kompenserende mål og martingaler

B.4.1 **Definition** Lad π være en sandsynlighed på $(\overline{\mathbb{R}}_0, \overline{\mathbb{B}}_0)$ med $\pi(\{0\}) = 0$. *Hazardmålet* ν for π er da målet på $(\mathbb{R}_0, \mathbb{B}_0)$ med tæthed

$$\frac{d\nu}{d\pi}(t) = \begin{cases} \frac{1}{\pi([t,\infty])} & \text{hvis } \pi([t,\infty]) > 0 \\ 0 & \text{ellers} \end{cases}$$

mht. restriktionen af π til $\mathbb{R}_0$.

B.4.2 **Definition** Hvis F er fordelingsfunktionen for en fordeling på $\overline{\mathbb{R}}_0$, så defineres *overlevelsesfunktionen* $\overline{F}$ ved $\overline{F} = 1 - F$.

B.4.3 Hvis π er en kontinuert fordeling på $\mathbb{R}_+$ med fordelingsfunktion F og tæthed f med hensyn til Lebesguemålet l, så har hazardmålet ν for π tæthed $\alpha(t) = \frac{f(t)}{1-F(t)}$ mht. Lebesguemålet l. Endvidere er $\overline{F}(t) = \exp(-\int_0^t \alpha(u)\,du)$. Vi kalder α *hazardfunktionen* for fordelingen π. Vi minder om, at når π er kontinuert, så er $\int_0^\infty \overline{F}(u)\,du = \mathrm{E}\,F$.

B.4.4 *Notation* Lad $P^{(0)}$ og $P^{(n)}_{t_1\ldots t_n}$, $n \in \mathbb{N}$, være markovkernerne, der bestemmer fordelingen af den simple punktproces T. Da er $\nu^{(n)}_{t_1\ldots t_n}$ hazardmålet for $P^{(n)}_{t_1\ldots t_n}$.

B.4.5 **Definition** *Kompensatoren* for fordelingen Q° på $(W, \mathcal{H}^\circ)$ af en simpel punktproces er processen Λ° defineret på $(W, \mathcal{H}^\circ)$ ved

$$\Lambda_t^\circ = \sum_{k=0}^{N_t^\circ - 1} \nu^{(k)}_{T_1^\circ\ldots T_k^\circ}(]T_k^\circ, T_{k+1}^\circ]) + \nu^{(N_t^\circ)}_{T_1^\circ\ldots T_{N_t^\circ}^\circ}(]T_{N_t^\circ}^\circ, t])$$

B.4.6 **Sætning** *Hvis Q° er en fordeling på $(W, \mathcal{H}^\circ)$ og Λ° dens kompensator, da gælder*

- Λ° *er voksende og højrekontinuert, specielt cadlag.*
- Λ° *er forudsigelig mht. filtreringen $(\mathcal{H}_t^\circ)$.*
- *Hvis $\mathrm{E}\, N_t^\circ < \infty$ for alle t, så er også $\mathrm{E}\,\Lambda_t^\circ < \infty$ for alle t. Endvidere er processen $M^\circ := N^\circ - \Lambda^\circ$ en Q°-martingal. Dvs. $(M_t^\circ, \mathcal{H}_t^\circ)$ er en martingal mht. Q°.*
- *Hvis $\mathrm{E}\, N_t^\circ < \infty$ for alle t og Λ° er kontinuert, så er $\mathrm{E}(M_t^\circ)^2 < \infty$ for alle t og $((M_t^\circ)^2 - \Lambda_t^\circ, \mathcal{H}_t^\circ)$ er en Q°-martingal.*
- *Hvis $\mathrm{E}\, N_t^\circ < \infty$ for alle t, så er Λ° op til Q°-uskelnelighed den eneste højrekontinuerte forudsigelige proces A° så $A_0^\circ = 0$ Q°-ns og $N^\circ - A^\circ$ er en Q°-martingal.*

B.4.7 *Notation* Med μ en mærket punktproces defineres den tilhørende tælleproces $\overline{N}$ ved $\overline{N}_t = N_t(E) = \mu([0,t] \times E)$.

B.4.8 *Notation* Lad $\nu^{(n)}_{t_1 \ldots t_n y_1 \ldots y_n}$ være hazardmålet for markovkernerne $P^{(n)}_{t_1 \ldots t_n y_1 \ldots y_n}$, der bestemmer fordelingen af tidspunkterne i en mærket punktproces.

B.4.9 **Definition** Lad Q° være en sandsynlighed på $(M, \sigma(M))$. Definer processerne $\overline{\Lambda}^\circ$ og $\Lambda^\circ(A)$, $A \in \mathcal{E}$, på $(M, \sigma(M))$ ved

$$\overline{\Lambda}_t^\circ = \sum_{k=0}^{\overline{N}_t^\circ - 1} \nu^{(k)}_{T_1^\circ \ldots T_k^\circ Y_1^\circ \ldots Y_k^\circ}(]T_k^\circ, T_{k+1}^\circ]) + \nu^{(\overline{N}_t^\circ)}_{T_1^\circ \ldots T^\circ_{\overline{N}_t^\circ} Y_1^\circ \ldots Y^\circ_{\overline{N}_t^\circ}}(]T^\circ_{\overline{N}_t^\circ}, t])$$

henholdsvis

$$\Lambda_t^\circ(A) = \int_{[0,t]} \pi^{(\overline{N}_{s-}^\circ)}_{T_1^\circ \ldots T^\circ_{\overline{N}_{s-}^\circ}, s, Y_1^\circ \ldots Y^\circ_{\overline{N}_{s-}^\circ}}(A)\, d\overline{\Lambda}_s^\circ$$

B.4.10 **Definition** Det *kompenserende mål* for fordelingen Q° af en mærket punktproces på $(M, \sigma(M))$ er det positive tilfældige mål L° defineret på $(M, \sigma(M))$ ved

$$L^\circ([0,t] \times A) = \Lambda_t^\circ(A)$$

for $t \in \mathbb{R}_0$ og $A \in \mathcal{E}$.

B.4.11 *Bemærkning* Ovenfor er diskuteret kompensatorer hhv. kompenserende mål på de kanoniske rum $(W, \mathcal{H}^\circ)$ og $(M, \sigma(M))$. Vi kan definere en kompensator for en tælleproces N defineret på $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ som den forudsigelige proces Λ, der gør $M := N - \Lambda$ til en P-martingal. Det er klart at denne kompensator Λ vil afhænge af filtreringen $(\mathcal{F}_t)$.

Hvis vi for tælleprocessen N har kompensatoren Λ mht. filtreringen $(\mathcal{F}_t)$ og ønsker at skifte filtrering til $(\mathcal{G}_t)$, så vil Λ også være kompensatoren mht. den nye filtrering, hvis

- $\mathcal{F}_t \subseteq \mathcal{G}_t$ for alle t, og $N - \Lambda$ er en martingal mht. P og filtreringen $(\mathcal{G}_t)$, da forudsigelighed bevares ved skift til en større filtrering.
- $\mathcal{G}_t \subseteq \mathcal{F}_t$ for alle t, og Λ er forudsigelig mht. $(\mathcal{G}_t)$, da martingalegenskaberne bevares ved skift til en mindre filtrering.

B.4.12 **Definition** Vi vil kalde λ *intensitetsprocessen* for en sandsynlighed Q° på $(W, \mathcal{H}^\circ)$ med kompensator Λ°, hvis λ er forudsigelig og $\Lambda^\circ_t = \int_0^t \lambda_s \, ds$ Q°-ns.

Tilsvarende vil vi kalde en samling $\lambda = (\lambda(A))_{A \in \mathcal{E}}$ af forudsigelige processer for en *intensitetsproces* for sandsynligheden Q° på $(M, \sigma(M))$, hvis Q°-ns. $A \mapsto \lambda_t(A)$ er et positivt mål på $(E, \mathcal{E})$ for alle t, og $L^\circ([0,t] \times A) = \int_0^t \lambda_s(A) \, ds$, $t \in \mathbb{R}_0$, $A \in \mathcal{E}$.

B.4.13 *Bemærkning* Hvis E er tællelig, så er det nok at specificere $\lambda(A)$ for etpunktsmængder $A = \{a\}$, $a \in E$. Vi skriver da $\lambda^a_t = \lambda_t(\{a\})$.

B.4.14 **Sætning** *Lad Q° være en sandsynlighed på $(M, \sigma(M))$ og $P^{(n)}_{t_1 \ldots t_n y_1 \ldots y_n}$, $\pi^{(n)}_{t_1 \ldots t_n, t, y_1 \ldots y_n}$ de tilhørende markovkerner. Hvis alle $P^{(n)}_{t_1 \ldots t_n, t, y_1 \ldots y_n}$ er absolut kontinuerte mht. Lebesguemålet med hazardfunktion $\alpha^{(n)}_{t_1 \ldots t_n y_1 \ldots y_n}$, så har Q° intensitetsproces λ med*

$$\lambda_t(A) = \alpha^{(\overline{N}^\circ_{t-})}_{T^\circ_1 \ldots T^\circ_{\overline{N}^\circ_{t-}} Y^\circ_1 \ldots Y^\circ_{\overline{N}^\circ_{t-}}}(t) \pi^{(\overline{N}^\circ_{t-})}_{T^\circ_1 \ldots T^\circ_{\overline{N}^\circ_{t-}}, t, Y^\circ_1 \ldots Y^\circ_{\overline{N}^\circ_{t-}}}(A)$$

B.5 Multiplikative intensitetsmodeller — Nelson-Aalen estimatoren

B.5.1 **Definition** Vi har en *multiplikativ intensitetsmodel* for tælleprocessen N, hvis N har en intensitetsproces $\lambda(t) = \alpha(t) Y(t)$ mht. en filtrering $(\mathcal{F}_t)$ og et sandsynlighedsmål P, hvor $\alpha(\cdot)$ er en deterministisk funktion og Y er en forudsigelig stokastisk proces. Situationen er, at α er ukendt, men Y er en observerbar proces.

B.5.2 Vi vil i denne ramme interessere os for at estimere integralet af α, $t \mapsto \int_0^t \alpha(u) \, du$.

B.5.3 *Eksempel* Lad π være en kontinuert levetidsfordeling på $\mathbb{R}_0$ med overlevelsesfunktion $\overline{F}$, tæthed f og hazardfunktion $\alpha = f / \overline{F}$.

Lad X være en stokastisk variabel med fordeling π. Så er $N_t = 1(X \leqq t)$ en tælleproces med intensitetsproces $\lambda(t) = \alpha(t) 1(t \leqq X)$. Vi har således en multiplikativ intensitet mht. filtreringen $(\mathcal{F}^N_t)$ frembragt af N.

B.5.4 I det følgende beskæftiger vi os med trunkering og censur. Trunkering angiver en betingning. Ved venstretrunkering betinger vi med at en levetid X er større end t_1, mens vi ved højretrunkering betinger med at X er mindre end t_2. Ved censur ser vi ikke X, men kun at X ligger i en given mængde. Ved højrecensur til tid t_2 ser vi, at $X > t_2$ og ved venstrecensur i t_1 ser vi $X < t_1$.

B.5.5 (Højrecensur) Højrecensur er behandlet i [ABGK93, afs. III.2]. Lad τ være en stoppetid mht. $(\mathcal{F}_t)$, og definer $C(t) = 1\,(t \leqq \tau)$. Så er C forudsigelig. Vi vil stoppe observation af tælleprocessen N med intensitet λ mht. $(\mathcal{F}_t)$ til tid τ. Antag $\mathrm{E}\, N_t < \infty$ for alle t. Vi har, at $N(\cdot) = \Lambda(\cdot) + M(\cdot)$, hvor M er en martingal mht. $(\mathcal{F}_t)$. Vi ser $N^c(\cdot) = \int_0^{\cdot} C(s)\, dN(s) = N(t \wedge \tau)$. Nu er $N^c(t) = \int_0^t C(s)\, d\Lambda(s) + \int_0^t C(s)\, dM(s)$, hvor sidste led er et stokastisk integral af en forudsigelig proces; igen en martingal mht. $(\mathcal{F}_t)$. Kompensatoren for N^c mht. $(\mathcal{F}_t)$ er således $\int_0^{\cdot} C(s)\, d\Lambda(s)$, og intensitetsprocessen er $\lambda^c(t) = \lambda(t)C(t)$.

Vi kalder det *uafhængig højrecensur*, når τ er en stoppetid mht. en større filtrering $(\mathcal{G}_t)$, og N har samme kompensator mht. $(\mathcal{G}_t)$ som $(\mathcal{F}_t)$.

B.5.6 (Venstretrunkering) Venstretrunkering er behandlet i [ABGK93, afsn. III.3]. Vi betragter kun situationen fra eksempel B.5.3. Vi betinger med, at $X > t_1$, hvor t_1 er et fast tidspunkt. Man indser nemt, at i den betingede fordeling af X givet $X > t_1$, er hazarden $\alpha(a)$, $a \geqq t_1$. (Se eventuelt [Jac95a, p. 4.5].) Tælleprocessen $\tilde{N}$, hvor vi har betinget, får derfor intensitet $\tilde{\lambda}(t) = \alpha(t)1\,(t \leqq X)\,1\,(t_1 < t)$ mht. filtreringen induceret af den oprindelige proces N.

B.5.7 *Eksempel* Man vil mere generelt få den multiplikative intensitetsstruktur, hvis man betragter iid levetider og tælleprocessen N, der tæller antal dødsfald vi har set. I disse tilfælde bliver den deterministiske proces altid hazarden α for levetiden og Y processen, der til tid t tæller antal individer, der er under observation umiddelbart før tid t.

Endvidere vil man nemt kunne håndtere venstretrunkering og højrecensur i denne ramme. Hvis vi i eksempel B.5.3 vil venstretrunkere levetiden i alder t_1 og højrecensurere i alder t_2, $t_1 < t_2$, så viser de foregående eksempler, at tælleprocessen $N^c(t) = N(t \wedge t_2) - N(t \wedge t_1) = \int_0^t 1\,(t_1 < s \leqq t_2)\, dN_s$ får intensitetsproces

$$\lambda^c(t) = \alpha(t)\underbrace{1\,(t \leqq X)\,1\,(t_1 < t \leqq t_2)}_{Y(t)}$$

mht. den oprindelige filtrering.

Har vi uafhængige levetider X_i, $i = 1, \ldots, n$, der trunkeres og censureres i aldre, der må afhænge af i, får vi uafhængige tælleprocesser N_i med intensitetsprocesser $\lambda_i(t) = \alpha(t)Y_i(t)$ mht. $(\mathcal{F}_t^{N_i})$. Da tælleprocesserne er uafhængige, giver [Jac95a, sætn. 6.1], at intensiteterne bevares, hvis vi betragter filtreringen frembragt af alle N_i'erne. Summen $\overline{N}(t) = \sum_{i=1}^n N_i(t)$ får intensitetsproces $\overline{\lambda}(t) = \sum_{i=1}^n \lambda_i(t) = \alpha(t)\sum_{i=1}^n Y_i(t)$ mht. denne filtrering. Her tæller $\sum_{i=1}^n Y_i(t)$ antal individer under risiko umiddelbart før tid t.

B.5.8 **Definition** Nelson-Aalen estimatoren for den integrerede intensitet $\beta(t) = \int_0^t \alpha(u)\, du$ er defineret ved

$$\hat{\beta}(t) = \int_0^t \frac{1\,(Y(s) > 0)}{Y(s)}\, dN(s) = \sum_{n=1}^{N(t)} \frac{1}{Y(T_n)}$$

B.5.9 *Bemærkning* Estimatoren motiveres f.eks. ved at $N - \alpha Y$ er en martingal, så $dN(t) = \alpha(t)Y(t)\,dt$ hvis vi ser bort fra "martingalstøjen". Dvs. $\hat{\alpha}(t)\,dt = \frac{dN(t)}{Y(t)}$.

B.5.10 *Bemærkning* Man kan også give estimatoren mening som en ikke-parametrisk maksimum likelihood estimator; se [ABGK93, afsn. IV.1.5].

Ved en ikke-parametrisk maksimumlikelihood estimator forstår vi normalt en fordeling, der "maksimerer likelihoodfunktionen". Dette skal forstås på den måde, at hvis vi kan finde en fordeling, der giver vores data positiv sandsynlighed, så vil vi foretrække den frem for en fordeling, der ikke giver vores data positiv sandsynlighed. Vi vil altså estimere vores fordeling med en fordeling, der har punktsandsynligheder i vores data. Man kan nu overbevise sig om, at en diskret fordeling, med punktsandsynligheder i lige netop de punkter, hvor vores data er, vil være den bedste. Tilbage er at finde disse punktsandsynligheder.

Der er forskellige problemer i denne metode. F.eks. skal vi udvide vores model. Den oprindelig model indeholder kun kontinuerte levetidsfordelinger, men vi ønsker at estimere med en diskret fordeling. Denne udvidelse er ikke entydig. I [Jac82, afsn. 4.6] er gennemgået én måde at foretage denne udvidelse på. Endvidere kan den fordeling vi vil estimere indgå i likelihoodfunktionen gennem en funktional, f.eks. $\frac{\overline{F}(x)}{\mathrm{E}\,F}$, så det ikke er punktsandsynlighederne direkte, der står i likelihoodfunktionen. Eksempler på disse problemer ses i [Var82a] og [MD95].

B.5.11 Vi vil i de øvrige kapitler bruge vendinger som "få en pæn struktur af intensiteten" og "estimere som vi plejer". Vi tænker da på, at vi kan estimere ved at bruge Nelson-Aalen estimatoren. Vi vil være specielt glade, hvis den er motiveret ved en multiplikativ intensitetsmodel, og ikke blot som en ikke-parametrisk maksimum likelihood estimator.

B.6 Likelihoodprocessen

B.6.1 *Notation* Lad Q°, $\tilde{Q}^\circ$ være to sandsynligheder på $(W, \mathcal{H}^\circ)$. Så betegner Q_t° og $\tilde{Q}_t^\circ$ restriktionerne til $\mathcal{H}_t^\circ$.

B.6.2 **Definition** $\tilde{Q}^\circ$ er *lokalt absolut kontinuert* mht. Q° hvis $\tilde{Q}_t^\circ \ll Q_t^\circ$ for alle $t \in \mathbb{R}_0$. I bekræftende fald skriver vi $\tilde{Q}^\circ \ll_{\mathrm{loc}} Q^\circ$ og definerer *likelihoodprocessen* $L = (L_t)_{t \geqq 0}$ ved $L_t = \frac{d\tilde{Q}_t^\circ}{dQ_t^\circ}$.

B.6.3 **Sætning** *Likelihoodprocessen L er en $(Q^\circ, \mathcal{H}_t^\circ)$-martingal, der har en cadlag version.*

B.6.4 **Definition** Om to sandsynligheder π, $\tilde{\pi}$ på $\overline{\mathbb{R}}_+$ siges, at $\tilde{\pi}$ er *lokalt absolut kontinuert* mht. π, og vi skriver $\tilde{\pi} \ll_{\mathrm{loc}} \pi$, hvis der om restriktionerne til $\mathbb{R}_+$ gælder $\tilde{\pi}\,|_{\mathbb{R}_+} \ll \pi\,|_{\mathbb{R}_+}$, og $\pi(]t, \infty]) > 0$ for alle t hvis $\tilde{\pi}(\{\infty\}) > 0$.

B.6.5 **Sætning** *Lad* Q°, $\tilde{Q}^\circ$ *være to sandsynligheder på* $(W, \mathcal{H}^\circ)$ *og* $P^{(n)}$, $\tilde{P}^{(n)}$ *de tilhørende markovkerner. Hvis for alle* $n \in \mathbb{N}_0$, $\tilde{P}^{(n)}_{t_1 \ldots t_n} \ll_{loc} P^{(n)}_{t_1 \ldots t_n}$, *så er* $\tilde{Q}^\circ \ll_{loc} Q^\circ$. *Hvis yderligere* $\tilde{P}^{(n)}_{t_1 \ldots t_n}$ *og* $P^{(n)}_{t_1 \ldots t_n}$ *er absolut kontinuerte og har hazardfunktioner* $\tilde{\alpha}^{(n)}_{t_1 \ldots t_n}$ *og* $\alpha^{(n)}_{t_1 \ldots t_n}$, *så er*

$$L_t = \exp(-(\tilde{\Lambda}^\circ_t - \Lambda^\circ_t)) \prod_{n=1}^{N^\circ_t} \frac{\tilde{\lambda}_{T^\circ_n}}{\lambda_{T^\circ_n}}$$

hvor $\tilde{\Lambda}^\circ$ *og* Λ° *er kompensatorerne for* $\tilde{Q}^\circ$ *og* Q°, *og* $\tilde{\lambda}$, λ *er intensitetsprocesserne* $\tilde{\lambda}_t = \tilde{\alpha}^{(\overline{N}^\circ_{t-})}_{T^\circ_1 \ldots T^\circ_{\overline{N}^\circ_{t-}}}(t)$ *og* $\lambda_t = \alpha^{(\overline{N}^\circ_{t-})}_{T^\circ_1 \ldots T^\circ_{\overline{N}^\circ_{t-}}}(t)$ *som i sætning B.4.14.*

Appendix C

Generelle punktprocesser

C.1 Introduktion

C.1.1 Dette appendix gengiver de resultater fra teorien om punktprocesser, som vi skal bruge. Resultaterne er primært taget fra [Jac96, p. 1–17]. Visse resultater har jeg ikke fundet andre steder og beviserne er derfor medtaget her.

C.2 Definition af en punktproces

C.2.1 Betragt det målelige rum $(S, \mathbb{S})$. Antag, at alle etpunktsmængder er målelige, dvs. $\{x\} \in \mathbb{S}$ for alle $x \in S$. Vi vil definere en punktproces på S.

C.2.2 **Definition** (Tællemål) Mængden af tællemål $\tilde{N}_S$ på S er mængden af mål μ° på $(S, \mathbb{S})$, der opfylder

1. $\mu^\circ(A) \in \overline{\mathbb{N}}_0$ for $A \in \mathbb{S}$
2. $\mu^\circ(A) < \infty$ for $A \in \mathbb{S}$ begrænset

Hvis yderligere $\mu^\circ(\{x\}) \in \{0, 1\}$ for $x \in S$ ligger μ° i mængden N_S.

C.2.3 *Bemærkning* Vi taler ovenfor om begrænsede mængder af S. Det kan vi ikke gøre uden at lægge mere struktur på S. Vi vil ikke gå i detaljer her, men blot bemærke, at vi i vores anvendelser vil have $S = \mathbb{R}^d$, hvor mængder af formen $A = [-N, N]^d$ for $N \in \mathbb{N}$ kan bruges, eller $S = \mathbb{R}^d \times \hat{S}$, hvor mængder af formen $A = [-N, N]^d \times \hat{S}$ opfylder kravet. Tilsvarende taler vi på et tidspunkt om kompakte mængder.

C.2.4 **Definition** Rummet N_S af tællemål udstyres med σ-algebraen $\sigma(N_S)$ initialt frembragt af familien af afbildninger $(\mu^\circ \mapsto \mu^\circ(A))_{A \in \mathbb{S}}$ fra N_S til $\overline{\mathbb{N}}_0$. Tilsvarende for $\tilde{N}_S$.

C.2.5 **Definition** En *simpel punktproces* er en stokastisk variabel μ defineret på $(\Omega, \mathcal{F}, P)$ med værdier i $(N_S, \sigma(N_S))$.

En *punktproces med multiple punkter* er en stokastisk variabel μ defineret på $(\Omega, \mathcal{F}, P)$ med værdier i $(\tilde{N}_S, \sigma(\tilde{N}_S))$.

Fordelingen af μ er den transformerede sandsynlighed $\mu(P)$ på $(N_S, \sigma(N_S))$ hhv. $(\tilde{N}_S, \sigma(\tilde{N}_S))$.

C.2.6 For $\mu^\circ \in N_S$ vil vi tænke på μ° som en mængde $(x_i)_{i\in I}$ af forskellige punkter i S. For $\mu^\circ \in \tilde{N}_S$ kan $x_i = x_j$ forekomme for $i \neq j$. Med $A \in \mathbb{S}$ er $\mu^\circ(A)$ antallet af punkter i mængden A. Vi kan derfor skrive $\mu^\circ(A) = \sum_{i\in I} 1\,(x_i \in A)$ eller $\mu^\circ = \sum_{i\in I} \varepsilon_{x_i}$.

C.2.7 *Bemærkning* Betingelsen $\mu^\circ(A) < \infty$ for $A \in \mathbb{S}$ begrænset sikrer, at punkterne $(x_i)_{i\in I}$ ikke har et fortætningspunkt. I vores anvendelser kan vi skrive S som en tællelig forening af begrænsede mængder, dvs. indexmængden I er højst numerabel.

C.3 Laplacefunktionaler

C.3.1 *Notation* Med ν et mål på S og $f : S \to \mathbb{R}_0$ en målelig funktion er $\nu(f) = \int f\, d\nu$. Hvis $\mu^\circ \in N_S$ er $\mu^\circ(f) = \int f\, d\mu^\circ = \sum_{i\in I} f(x_i)$.

C.3.2 **Definition** Intensitetsmålet Λ for punktprocessen μ defineres ved $\Lambda(A) = \mathrm{E}\,\mu(A)$, $A \in \mathbb{S}$. Hvis intensitetsmålet Λ har tæthed λ mht. et mål π på S, så vil vi kalde λ *π-intensiteten*.

C.3.3 **Definition** *Laplacefunktionalen* L_μ for en punktproces μ på $(S, \mathbb{S})$ defineres som afbildningen $f \mapsto L_\mu(f)$ fra $b_+\mathbb{S}$ (rummet af målelige, positive, begrænsede $f : S \to \mathbb{R}_0$) til $\mathbb{R}_0$ givet ved $L_\mu(f) = \mathrm{E}\, e^{-\mu(f)}$.

C.3.4 **Sætning** *Fordelingen af en punktproces μ er entydigt bestemt af Laplacefunktionalen L_μ. Det er nok at betragte stikprøvefunktioner f med kompakt støtte.*

C.4 Poissonprocesser

C.4.1 **Definition** En simpel punktproces μ er en *poissonproces*, hvis

1. $\mu(A)$ er poissonfordelt for alle $A \in \mathbb{S}$, A begrænset.
2. $\mu(A_1), \ldots, \mu(A_n)$ er stokastisk uafhængige for $n \geqq 2$, $A_1, \ldots, A_n \in \mathbb{S}$ parvis disjunkte.

C.4.2 *Bemærkning* Hvis μ er simpel punktproces, der opfylder 1 i definition C.4.1 så vil μ automatisk opfylde punkt 2. Beviset er dybt og følger af [Jac96, "Chok" p. 4].

C.4.3 *Bemærkning* (Konstruktion af poissonprocesser) En poissonproces μ på $A \in \mathbb{S}$ med $\Lambda(A) < \infty$ konstrueres, ved at lade $N, X_1, X_2, \dots$ være uafhængige, N poissonfordelt $\Lambda(A)$ og X_i identisk fordelte på A med fordeling $\pi(\cdot) = \frac{\Lambda(\cdot)}{\Lambda(A)}$. Så er $\mu = \sum_{i=1}^{N} \varepsilon_{X_i}$ den ønskede poissonproces. En poissonproces på hele S, med $\Lambda(S) = \infty$ mulig, konstrueres ved at sammensætte områder A_i med $\Lambda(A_i) < \infty$.

C.4.4 *Bemærkning* (Kombination af uafhængige poissonprocesser) Lad $\mu_1 = \sum \varepsilon_{x_i}$ og $\mu_2 = \sum \varepsilon_{\tilde{x}_j}$ være uafhængige poissonprocesser på $(S_1, \mathbb{S}_1)$ og $(S_2, \mathbb{S}_2)$ med intensitetsmål Λ_1 og Λ_2. Dette svarer oplagt til én poissonproces på den disjunkte forening[1] $S_1 \dot{\cup} S_2$.

C.4.5 *Bemærkning* Lad μ være en poissonproces med intensitetsmål Λ og $A \in \mathbb{S}$ med $0 < \Lambda(A) < \infty$. Givet $\mu(A) = n$ er punkterne $X_1, \dots, X_n$ i mængden A iid med fordeling $\pi(\cdot) = \frac{\Lambda(\cdot)}{\Lambda(A)} 1\,(\cdot \in A)$.

De n punkter i S er indiceret med tallene $1, \dots, n$. Vi skal imidlertid angive en fremgangsmåde at indicere punkterne på, da det ellers ikke generelt vil gælde. Vi vil tildele vores punkter indicis tilfældigt. Dvs. vi nummererer punkterne efter en ligefordeling på de $n!$ måder de n punkter kan nummereres på. Disse problemer er f.eks. omtalt i [DVJ88, afsn. 5.3], hvor vi betragter vores observation modulo permutationer af index. Vi vil udvide det lidt:

Lad μ_k være en poissonproces på S_k med intensitetsmål Λ_k, $k = 1, 2$. Antag μ_1 og μ_2 er uafhængige. Lad $A_1 \in \mathbb{S}_1$, $A_2 \in \mathbb{S}_2$ med $0 < \Lambda_1(A_1) < \infty$ og $0 < \Lambda_2(A_2) < \infty$. Givet $\mu_1(A_1) + \mu_2(A_2) = n$ er punkterne $X_1, \dots, X_n$ iid med fordeling

$$\begin{aligned}
P(X_i \in A_1) = 1 - P(X_i \in A_2) &= \frac{\Lambda_1(A_1)}{\Lambda_1(A_1) + \Lambda_2(A_2)} \\
P(X_i \in \cdot\, | X_i \in A_1) &= \frac{\Lambda_1(\cdot)}{\Lambda_1(A_1)} \\
P(X_i \in \cdot\, | X_i \in A_2) &= \frac{\Lambda_2(\cdot)}{\Lambda_2(A_2)}
\end{aligned}$$

Bidraget fra observationen $X_i = x_i$ til likelihoodfunktionen er

$$\frac{d\Lambda_1(x_i)^{1(x_i \in A_1)}\, d\Lambda_2(x_i)^{1(x_i \in A_2)}}{\Lambda_1(A_1) + \Lambda_2(A_2)}$$

C.4.6 **Sætning** *Hvis μ er en poissonproces på $(S, \mathbb{S})$ med intensitetsmål Λ og $A, B \in \mathbb{S}$ med $B \subseteq A$ og $0 < \Lambda(B) \leqq \Lambda(A) < \infty$, så gælder:*

Hvis vi observerer μ over A og betinger med antallet af punkter $\mu(A) = n$, og derefter betinger med, at de n punkter ligger i B, så følger punkterne samme fordeling, som hvis vi observerede μ over B og betingede med antallet af punkter i B.

[1] $S_1 \dot{\cup} S_2 = \{(i, s) | (i = 1 \wedge s \in S_1) \vee (i = 2 \wedge s \in S_2)\}$.

C.4.7 **Sætning** *Laplacefunktionalen* L_μ *for en poissonproces* μ *med intensitetsmål* Λ *er* $L_\mu(f) = \exp(-\Lambda(1-e^{-f}))$, *hvor* $\Lambda(1-e^{-f}) = \int_S (1-e^{-f(x)})\, d\Lambda(x)$.

C.4.8 **Sætning** (Likelihoodfunktion for poissonprocesser) *Lad* $(\Lambda_\theta)_{\theta\in\Theta}$ *være en familie af intensitetsmål på* S *og* Λ' *et andet intensitetsmål på* S. *Hvis vi observerer poissonprocessen* μ *med fordeling* P_θ *over en mængde* $A \in \mathbb{S}$, *der opfylder* $\Lambda'(A) < \infty$ *og for alle* $\theta \in \Theta$ *er* $\Lambda_\theta(A) < \infty$ *og* $\Lambda_\theta\mid_A \ll \Lambda'\mid_A$, *så er* $P_\theta\mid_A \ll P'\mid_A$ *og*

$$\frac{dP_\theta\mid_A}{dP'\mid_A}(\mu) = \exp\left(-(\Lambda_\theta(A) - \Lambda'(A))\right) \prod_{i=1}^{\mu(A)} \frac{d\Lambda_\theta\mid_A}{d\Lambda'\mid_A}(x_i) \tag{C.1}$$

hvor $x_1, \dots, x_{\mu(A)}$ *er de af punkterne fra* μ, *der ligger i* A.

C.4.9 **Korollar** *Hvis betingelserne fra sætning C.4.8 er opfyldt og vi yderligere har intensiteter,* $\Lambda_\theta = \lambda_\theta \cdot \pi$ *og* $\Lambda' = \lambda' \cdot \pi$, *så er*

$$\frac{dP_\theta\mid_A}{dP'\mid_A}(\mu) = \exp\left(-\int_A \lambda_\theta(x) - \lambda'(x)\, d\pi(x)\right) \prod_{i=1}^{\mu(A)} \frac{\lambda_\theta(x_i)}{\lambda'(x_i)}$$

C.4.10 **Definition** Lad $(\tilde{S}, \tilde{\mathbb{S}})$ være et måleligt rum. Definer funktionsrummet $\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}$, som rummet af funktioner $\mu : \tilde{\mathbb{S}} \to \overline{\mathbb{R}}_0$. Udstyr $\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}$ med σ-algebraen $\sigma(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}})$ induceret af projektionerne $\mu \mapsto \mu(A)$ for $A \in \tilde{\mathbb{S}}$.

Lad μ og $\tilde{\mu}$ være to stokastiske variable defineret på $(\Omega, \mathcal{F}, P)$ med værdier i $\left(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}, \sigma(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}})\right)$. Vi siger, at $\tilde{\mu}$ er en *version* af μ, hvis der for alle $A \in \tilde{\mathbb{S}}$ gælder

$$P(\tilde{\mu}(A) = \mu(A)) = 1$$

C.4.11 *Bemærkning* Generelt gælder ikke $N_{\tilde{S}} \in \sigma(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}})$.

C.4.12 **Sætning** (Transformation af poissonproces) *Lad* $(S, \mathbb{S})$ *og* $(\tilde{S}, \tilde{\mathbb{S}})$ *være målelige rum. Betragt en afbildning* $M : \tilde{\mathbb{S}} \to \mathbb{S}$, *og lad* μ *være en poissonproces på* $(S, \mathbb{S})$ *med intensitetsmål* Λ. *Så er* $\tilde{\mu} = \mu \circ M$ *en stokastisk variabel med værdier i* $\left(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}}, \sigma(\overline{\mathbb{R}}_0^{\tilde{\mathbb{S}}})\right)$. *Hvis*

- S *er et Borelrum.*
- $M(\emptyset) = \emptyset$.
- M *er tælleligt additiv: for* $A_1, A_2, \ldots \in \tilde{\mathbb{S}}$ *er* $M(A_1 \cup A_2 \cup \cdots) = M(A_1) \cup M(A_2) \cup \cdots$
- *For* $a \in \tilde{S}$ *er* $\Lambda(M(\{a\})) = 0$.
- *For* $A \in \tilde{\mathbb{S}}$ *begrænset er* $\Lambda(M(A)) < \infty$.

- *For $A_1, A_2 \in \tilde{\mathbb{S}}$ med $A_1 \cap A_2 = \emptyset$ er $\Lambda(M(A_1) \cap M(A_2)) = 0$.*

så har $\tilde{\mu}$ en version $\hat{\mu}$, der er en poissonproces på $(\tilde{S}, \tilde{\mathbb{S}})$ med intensitetsmål $\tilde{\Lambda} = \Lambda \circ M$. Hvis yderligere

- *For $a \in \tilde{\mathbb{S}}$ er $M(\{a\}) = \emptyset$ eller en etpunktsmængde.*
- *M bevarer begrænsethed: for $A \in \tilde{\mathbb{S}}$ begrænset er $M(A)$ begrænset.*
- *M bevarer disjunkthed: for $A_1, A_2 \in \tilde{\mathbb{S}}$ med $A_1 \cap A_2 = \emptyset$ er $M(A_1) \cap M(A_2) = \emptyset$.*

kan vi udelade *betingelsen "S er et Borelrum", og $\tilde{\mu}$ vil selv være en poissonproces med intensitet $\tilde{\Lambda}$.*

Bevis Da μ er en simpel punktproces, bliver $\tilde{\mu}$ er en stokastisk variabel som påstået. Vi får, at $\tilde{\mu}(\emptyset) = \mu(\emptyset) = 0$. De to første krav giver, at M er endeligt additiv. For $A_1, A_2, \ldots \in \tilde{\mathbb{S}}$ parvis disjunkte er

$$\begin{aligned}
\tilde{\mu}\left(\cup_{n=1}^{\infty} A_n\right) &= \mu\left(M\left(\cup_{n=1}^{\infty} A_n\right)\right) \\
&= \mu\left(\cup_{n=1}^{\infty} M(A_n)\right) \\
&= \mu\left(\cup_{n=1}^{\infty}\left[M(A_n) \setminus \cup_{i=1}^{n-1} M(A_i)\right]\right) \\
&= \sum_{n=1}^{\infty} \mu\left(M(A_n) \setminus \cup_{i=1}^{n-1} M(A_i)\right) \\
&= \sum_{n=1}^{\infty} \mu\left(M(A_n) \setminus M(\cup_{i=1}^{n-1} A_i)\right) \\
&\overset{ns}{=} \sum_{n=1}^{\infty} \mu\left(M(A_n)\right)
\end{aligned}$$

da $\mu(M(A_n) \setminus M(\cup_{i=1}^{n-1} A_i)) \overset{ns}{=} \mu(M(A_n) \setminus M(\cup_{i=1}^{n-1} A_i)) + \mu(M(A_n) \cap M(\cup_{i=1}^{n-1} A_i)) = \mu(M(A_n))$ thi $\mu(M(A_n) \cap M(\cup_{i=1}^{n-1} A_i)) = 0$ ns, thi A_n og $\cup_{i=1}^{n-1} A_i$ er disjunkte og $\Lambda(M(A_n) \cap M(\cup_{i=1}^{n-1} A_i)) = 0$. Der er nu to situationer:

- Hvis M bevarer disjunkthed, gælder den sidste lighed overalt, og ikke kun ns, da $\mu(\omega)$ er et mål. Dette viser, at $\tilde{\mu}$ er et mål, og vi indser tilsvarende, at hvis de yderligere betingelser om M gælder, så er $\tilde{\mu}$ en simpel punktproces.
- Hvis kun de oprindelige betingelser om M gælder, ser vi ikke af disse regninger, at $\tilde{\mu}$ er et mål. Problemet er, at nulmængden afhænger af følgen $A_1, A_2, \ldots$ af mængder. Der vil typisk være overtællelig mange sådanne følger. Vi er i samme situation som i [Jac95b] kapitel 4 om betingede middelværdier og regulære betingede sandsynligheder. Vi har der et konstruktivt bevis for eksistensen af en regulær betinget fordeling af en reel stokastisk variabel givet en σ-algebra. Dette udvides til eksistensen af

en regulær betinget fordeling af en stokastisk variabel med værdier i et Borelrum givet en σ-algebra. Vi kan gå gennem samme konstruktion, til at få en version $\hat{\mu}$ af $\tilde{\mu}$, hvor $\hat{\mu}$ er et mål på $(\tilde{S}, \tilde{\mathbb{S}})$. Vi skal undervejs udnytte egenskaberne for M, til at konstruere $\hat{\mu}$, så $\hat{\mu}$ bliver en simpel punktproces. Vi går ikke i yderligere detaljer.

For $A \in \tilde{\mathbb{S}}$ begrænset er $\tilde{\mu}(A) = \mu(M(A))$ poissonfordelt med middelværdi $\Lambda(M(A))$. Dette viser, at $\hat{\mu}$ er en poissonproces med det angivne intensitetsmål. $\square$

C.4.13 *Bemærkning* I konkrete situationer kan vi ofte fjerne passende Λ nulmængder fra S, og derved skaffe os en ny mængde S_{-}, en poissonproces μ_{-} på S_{-} med intensitetsmål Λ_{-}, samt en afbildning M_{-} således, at M_{-} opfylder det udvidede sæt af krav, og vi stadig har beskrevet vores fænomen. Et eksempel på dette ses i fodnote 4 på side 39.

C.4.14 **Korollar** *Lad der være givet målelige rum $(S, \mathbb{S})$, $(\tilde{S}_i, \tilde{\mathbb{S}}_i)$, $i = 1, \ldots, n$, og afbildninger $M_i : \tilde{\mathbb{S}}_i \to \mathbb{S}$, $i = 1, \ldots, n$. Lad μ være en poissonproces på S med intensitetsmål Λ. Antag, at S er et Borelrum og, at afbildningerne M_i, $i = 1, \ldots, n$, opfylder det første sæt krav i sætningen. Hvis alle fællesmængderne $M_i(\tilde{S}_i) \cap M_j(\tilde{S}_j)$, $i \neq j$, har Λ mål 0, så har afbildningerne $\tilde{\mu}_i = \mu \circ M_i$, $i = 1, \ldots, n$, versioner $\hat{\mu}_i$, $i = 1, \ldots, n$, som er uafhængige poissonprocesser.*

Bevis Udvid beviset fra før. $\square$

C.4.15 **Korollar** *Lad μ være en poissonproces på $(S, \mathbb{S})$ med intensitetsmål Λ og $t : (S, \mathbb{S}) \to (\tilde{S}, \tilde{\mathbb{S}})$ en målelig afbildning. Hvis*

- *S er et borelrum.*
- *$t(\Lambda)(\tilde{A}) < \infty$ for $\tilde{A} \in \tilde{\mathbb{S}}$ begrænset.*
- *$t(\Lambda)(\{\tilde{a}\}) = 0$ for $\tilde{a} \in \tilde{S}$.*

så har $t(\mu)$ en version, der er en poissonproces på $(\tilde{S}, \tilde{\mathbb{S}})$ med intensitetsmål $t(\Lambda)$. Hvis t er injektiv og originalmængdeafbildningen t^{-1} bevarer begrænsethed, så kan vi udelade betingelsen "S er et borelrum", og $t(\mu)$ er selv en poissonproces.

C.4.16 **Lemma** *Hvis Λ har intensitet $\lambda \circ t$ mht. π, så har $t(\Lambda)$ intensitet λ mht. $t(\pi)$. Hvis specielt t er injektiv har vi $\lambda = (\lambda \circ t^{-1}) \circ t$ og intensiteten for $t(\Lambda)$ bliver $\lambda \circ t^{-1}$ mht. $t(\pi)$.*

Litteratur

Referencer markeret med Ξ har jeg ikke læst.

[ABE+55] M. Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman, *An emperical distribution function for sampling with incomplete information*, Annals of Mathematical Statistics **26** (1955), 641–647, Ξ.

[ABGK93] Per Kragh Andersen, Ørnulf Borgan, Richard D. Gill, and Niels Keiding, *Statistical Models Based on Counting Processes*, Springer Series in Statistics, Springer-Verlag, 1993.

[AJ91] Steen Arne Andersson and Søren Tolver Jensen, *Forelæsningsnoter i Sandsynlighedsregning*, 3. ed., Institut for Matematisk Statistik, Københavns Universitet, September 1991, Består af hæfte 1 og 2.

[Alh92] Juha M. Alho, *On Prevalence, Incidence, and Duration in General Stable Populations*, Biometrics **48** (1992), 587–592.

[Arj85] Elja Arjas, *Discussion of paper by P. K. Andersen and Ø. Borgan*, Scandinavian Journal of Statistics **12** (1985), 150–153.

[Arj86] Elja Arjas, *Stanford heart transplantation data revisited: a real time approach*, Modern Statistical Methods in Chronic Disease Epidemiology (S. H. Moolgavkar and R. L. Prentice, eds.), Wiley, New York, 1986, Ξ, pp. 65–81.

[BAL80] K. P. Burnham, D. R. Anderson, and J. L. Laake, *Estimation of density from line transect sampling of biological populations*, Wildlife Monographs, Journal of Wildlife Management **44** (1980), suppl., Ξ.

[BBBB72] R. E. Barlow, D. J. Bartholomew, J. M. Bremner, and H. D. Brunk, *Statistical Inference under Order Restrictions*, Wiley, New York, 1972, Ξ.

[Ber92] Christian Berg, *Kapitel II. Mål- og integralteori*, Matematik 2 MA, Matematisk Analyse, Matematisk Institut, Københavns Universitet, 1991–92, Består af del A og B.

[Bri86] David R. Brillinger, *The Natural Variability of Vital Rates and Associated Statistics*, Biometrics **42** (1986), 693–734, with discussion.

[Chi86] Chin Long Chiang, *Discussion on the paper by David R. Brillinger*, Biometrics **42** (1986), 714–717.

[DV86] L. Denby and Y. Vardi, *The survival curve with decreasing density*, Technometrics **28** (1986), 359–367, Ξ.

[DVJ88] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*, Springer Series in Statistics, Springer-Verlag, 1988.

[Fel71] William Feller, *An Introduction to Probability Theory and Its Applications*, 2. ed., vol. II, Wiley, 1971.

[Gil80] Richard D. Gill, *Nonparametric Estimation Based on Censored Observation of a Markov Renewal Process*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **53** (1980), 97–116.

[Gil83] Richard David Gill, *Discussion of the papers by Helland and Kurtz*, Bulletin of the International Statistical Institute, Proceedings of the 44$^{\text{th}}$ session, vol. 50, 1983, Book 3, pp. 239–243.

[Gre56] U. Grenander, *On the theory of mortality measurement, part II*, Skandinavisk Aktuarie Tidsskrift **39** (1956), 126–153, Ξ.

[Gro85] P. Groeneboom, *Estimating a monotone density*, Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (Hayward: Wadsworth) (L. M.L̃e Cam and R. A. Olshen, eds.), vol. 11, 1985, Ξ, pp. 539–555.

[Gro87] P. Groeneboom, *Asymptotics for incomplete censored observations*, Report 87–18, Department of Mathematics, University of Amsterdam, 1987, Ξ.

[Jac82] Martin Jacobsen, *Statistical Analysis of Counting Processes*, Lecture Notes in Statistics, vol. 12, Springer-Verlag, 1982.

[Jac95a] Martin Jacobsen, *Lecture notes on Marked Point Processes*, Institute of Mathematical Statistics, University of Copenhagen, August 1995.

[Jac95b] Martin Jacobsen, *Videregående sandsynlighedsregning*, Institut for Matematisk Statistik, Københavns Universitet, 1995.

[Jac96] Martin Jacobsen, *Generelle Punktprocesser — Resumé af forelæsninger*, Institut for Matematisk Statistik, Københavns Universitet, februar 1996.

[JK90] Martin Jacobsen and Niels Keiding, *Markovkæder*, 3. ed., Institut for Matematisk Statistik, Københavns Universitet, Juni 1990.

[Kar91] Allan F. Karr, *Point Processes and Their Statistical Inference*, 2. ed., Probability: Pure and Applied, Marcel Dekker, 1991.

[Kei86] Niels Keiding, *Discussion of the paper by David R. Brillinger*, Biometrics **42** (1986), 721–724.

[Kei90] Niels Keiding, *Statistical inference in the Lexis diagram*, Philosophical Transactions of the Royal Society London A **332** (1990), 487–509.

[Kei91] Niels Keiding, *Age-specific Incidence and Prevalence: a Statistical Perspective*, Journal of the Royal Statistical Society A **153** (1991), no. 3, 371–412, with discussion.

[Kei92] Niels Keiding, *Independent delayed entry*, Survival Analysis: State of the Art (J. P. Klein and P. K. Goel, eds.), NATO ASI Series, Series E: Applied Sciences, vol. 211, Kluwer Academic Publishers. Printed in the Netherlands, 1992, pp. 309–326.

[Lau83] Steffen L. Lauritzen, *Lectures on Statistical Inference*, Institute of Mathematical Statistics, University of Copenhagen, 1983.

[MD95] Sally McClean and Colum Devine, *A nonparametric maximum likelihood estimator for incomplete renewal data*, Biometrika **82** (1995), no. 4, 791–803.

[Rao69] B. L. S.P̃rakasa Rao, *Estimation of a unimodal density*, Sankhyã A **31** (1969), 23–36, Ξ.

[Sim80] Richard Simon, *Length biased sampling in etiologic studies*, American Journal of Epidemiology **111** (1980), no. 4, 444–452.

[Var82a] Yehuda Vardi, *Nonparametric estimation in the presence of length bias*, The Annals of Statistics **10** (1982), no. 2, 616–620.

[Var82b] Yehuda Vardi, *Nonparametric estimation in renewal processes*, The Annals of Statistics **10** (1982), no. 3, 772–785.

[Var88] Yehuda Vardi, *Statistical Models for Intercepted Data*, Journal of the American Statistical Association **83** (1988), no. 401, 183–197.

[Var89] Yehuda Vardi, *Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation*, Biometrika **76** (1989), no. 4, 751–761.

[VC84] Joseph G. Voelkel and John Crowley, *Nonparametric inference for a class of semi-markov processes with censored observations*, The Annals of Statistics **12** (1984), no. 1, 142–160.

[Weg70] E. J. Wegman, *Maximum likelihood estimation of a unimodal density function*, Annals of Mathematical Statistics **41** (1970), 457–471, Ξ.