

Over betrouwbaarheidsbanden en aanpassingstoetsen bij gecensureerde waarnemingen.

Otto J.W.F. Kardaun^(*)

Samenvatting

Bij levensduurstudies treedt vaak het verschijnsel op dat een aantal waarnemingen gecensureerd zijn. Dit betekent, dat men in die gevallen niet de precieze levensduur heeft kunnen bepalen (bijvoorbeeld doordat een individu op het einde van de studie nog in leven is), maar wel weet dat deze groter dan een bepaalde tijdsduur moet zijn. In een dergelijke situatie wordt de Kaplan-Meier (1958) schatter veel gebruikt om de overlevingsfunctie $S(t)$ (dit is de kans om nog tenminste een tijd t , na een of andere specifieke gebeurtenis, in leven te blijven) verdelingsvrij te schatten. Om het effect van de eindigheid van de steekproefomvang tot uitdrukking te brengen, is men er toe over gegaan betrouwbaarheidsbanden voor de onbekende overlevingsfunctie $S(t)$ te construeren.

In dit artikel, dat de uitwerking vormt van een voordracht gehouden tijdens de MBS workshop in Heeze, zal, mede in het licht van recente publicaties hierover, op verschillende typen betrouwbaarheidsbanden voor $S(t)$ worden ingegaan. Het verband met aanpassingstoetsen zal worden aangegeven en als motivatie van het geheel zal dienen de uitwerking van een aantal concrete levensduurgegevens betreffende stadium I larynxkanker patienten.

Inhoud

0. Inleiding.

1. Waarschijnlijkheidstheoretisch model.
2. Kaplan-Meier, Nelson en Greenwood.
3. Instantane betrouwbaarheidsbanden.
4. Simultane betrouwbaarheidsbanden (discreet).
5. Simultane betrouwbaarheidsbanden (continu).
6. Toepassing.

(*) Mathematisch Instituut, Rijksuniversiteit Groningen,
Hoogbouw W.S.N., k.348, 9700 AV Groningen, tel. 050-116795.

Inleiding.

We beschouwen, ter motivatie van de probleemstelling, de volgende gegevens:

0.6 1.3 2.4 2.5* 3.2 3.2* 3.3 3.3* 3.5 3.5 4.0 4.0 4.3 4.5* 4.5* 5.3 5.5*
5.9* 5.9* 6.0 6.1* 6.2* 6.4 6.5 6.5* 6.7* 7.0* 7.4 7.4* 8.1* 8.1* 9.6* 10.7*.

Deze getallen stellen de resterende levensduren voor (in jaren na behandeling) van 33 manlijke stadium I larynx-kanker patienten. De getallen zonder * geven precies (± 0.05 jaar) de tijdsduur aan tussen behandeling en de sterftedatum van de patient. Een * duidt aan dat de patient op de afsluitingsdatum van de studie (1-3-'81) nog in leven was. Het genoteerde getal is in dat geval de tijd tussen behandeling en 1-3-'81. De werkelijke resterende levensduur moet dan minstens zo groot zijn, maar is (nog) niet precies vast te stellen. Men spreekt in zo'n geval van een gecensureerde waarneming.

De leeftijd van de patienten op het moment van de behandeling is ook bekend. De verdeling van de leeftijd wordt samengevat door (minimum, 1^e kwartiel, mediaan, 3^e kwartiel, maximum) = (43,57,64,74,86). Bij het toepassen van het proportionele hazards model van Cox, zoals uitgewerkt in een uitvoeriger case-study, (KARDAUN (1983)) kon geen effect van de leeftijd worden aangetoond. De schatting voor de proportionaliteitsconstante was zodanig (0.90 met een 95% betrouwbaarheidsinterval van (0.55,1.50) voor het effect van 10 jaar hogere leeftijd), dat we ter vereenvoudiging van de situatie zullen aannemen dat de leeftijd ten tijde van behandeling (in een bepaalde range, zeg van 45 tot 80 jaar) gedurende de eerste 10 jaar geen invloed op de resterende levensduur van de patient heeft. Medische redenen voor een dergelijk gering effect van de leeftijd als hier waargenomen, zouden kunnen zijn dat de kanker krachtiger groeit bij jongere dan bij oudere mensen en verder dat de kanker nog een belangrijke doodsoorzaak is.

De volgende vragen doen zich voor. Zij geven de problemen aan waarmee we ons bij deze gelegenheid bezig zullen houden.

(1) Sterft dit type patienten sneller of niet sneller dan normale Nederlanders van een bepaalde leeftijd? (Het 'goodness-of-fit' probleem).

(2) Hoe kunnen voor de onbekende overlevingsfunctie $S(t)$ van de patienten betrouwbaarheidsbanden geconstrueerd worden, die de effecten van de eindige steekproefomvang weerspiegelen? (Het 'confidence band' probleem).

We zullen deze twee enigszins verbonden problemen vanuit 3 gezichtspunten benaderen.

A: Instantaan, voor één argument t van $S(t)$.

B: Simultaan (discreet), voor verscheidene waarden $t_1, \dots, t_k$ van $S(t)$.

C: Simultaan (continu), voor $S(t)$ op het interval $[0, T]$, waarbij T een geschikt gekozen eindtijd is.

Interessant is verder nog op te merken dat men bij de beantwoording van vraag (1) de leeftijd van de normale Nederlanders kan laten variëren, en op die manier een betrouwbaarheidsinterval voor de 'effectieve leeftijd' van de patienten (als gebied van de niet verworpen nulhypothese) kan construeren. Op dit laatste zullen we in de praktische toepassing nader ingaan.

Waarschijnlijkheidstheoretisch model.

Om het geheugen wat op te frissen volgen nu enkele veel gebruikte begrippen uit de overlevingsanalyse. We beschouwen toevallig verdeelde tijdsduren met (absoluut) continue verdelingsfunctie $F(t)$ gedefinieerd op $R^+ = [0, \infty)$, kansdichtheid $f(t) = F'(t)$, overlevingsfunctie $S(t) = 1 - F(t)$, hazard rate ('uitvalsnelheid') $\lambda(t) = f(t)/S(t) = -S'(t)/S(t)$ en cumulatieve hazard rate $\Lambda(t) = \int_0^t \lambda(u) du = -\ln S(t)$. Elk van deze 5 functies is in elk van de 4 andere uit te drukken. Het zal blijken dat sommige aspecten eenvoudiger zijn wanneer de theorie uitgedrukt wordt in $\Lambda(t)$ dan wanneer ze uitgedrukt wordt in $S(t)$. De hazard rate $\lambda(t)$ is de helling van de grafiek van $\Lambda(t)$ en heeft de interpretatie van conditionele kansdichtheid om op tijd t te sterven, gegeven dat men tot tijd t in leven is gebleven.

Hoewel er verschillende, meer en minder algemene, censureringsmodellen bestaan waarvoor theorie is ontwikkeld (zie bijv. GILL (1980)), zullen we ons hier beperken tot het model van random censurering, dat zich als volgt laat

beschrijven: De overlevingstijden $T_1^0, \dots, T_n^0$ (in willekeurige volgorde) worden beschouwd als onafhankelijke, identiek verdeelde toevallige grootheden met overlevingsfunctie $S(t)$ (dit wordt afgekort tot: $T_1^0, \dots, T_n^0 \sim S$). De censureringstijden (tijden tussen behandeling en einde van de studie) worden evenzo als onafhankelijke en identiek verdeelde toevallige grootheden opgevat: $U_1^0, \dots, U_n^0 \sim G$. De waarnemingen in de inleiding zijn de uitkomsten (in opklimmende volgorde) van (T_j, Δ_j) , $j=1, \dots, n$ waarbij $T_j = \min(T_j^0, U_j^0)$ en $\Delta_j = I_{\{T_j^0 < U_j^0\}}$ een indicator variabele is, die aangeeft of de waarneming gecensureerd is ($\Delta_j=0$; een $*$) of niet ($\Delta_j=1$; geen $*$). Er wordt ook aangenomen dat voor $j=1, \dots, n$ T_j^0 en U_j^0 onafhankelijk zijn, zodat $T_j \sim S(t)G(t)$. Het gaat erom om op grond van de waarnemingen uitspraken te doen over S , waarbij de functie G als 'nuisance parameter' wordt opgevat. We zullen bij de straks te bespreken asymptotische eigenschappen van schatters en betrouwbaarheidbanden niet steeds de voorwaarden vermelden waaronder een en ander geldt. Het is voldoende te eisen dat er sprake is van random censurering, terwijl parameter vrije schatters (voor S, G, Λ enz.) slechts beschouwd worden op een interval $[0, T]$, waarbij T een vaste eindtijd is zodat $S(T)G(T) > 0$ (BRESLOW & CROWLEY (1974)). Onder zwakke additionele voorwaarden zijn deze asymptotische eigenschappen, gebruik makend van martingaal technieken, ook bewezen voor algemenere censureringsmodellen. Men zie het werk van ANDERSON, BORGAN, GILL & KEIDING (1982).

Kaplan-Meier, Nelson en Greenwood.

Als laatste voorbereiding zullen we enige parameter vrije schatters voor S en Λ in ogenschouw nemen. Zij $\tau_1 < \tau_2 < \dots < \tau_m$ de rij van waargenomen sterftetijden gerangschikt in opklimmende volgorde. In het voorbeeld dus $0.6 < 1.3 < 2.4 < 3.2 < 3.3 \dots < 7.4$. De Kaplan-Meier schatter voor de overlevingsfunctie S wordt gegeven door

$$\hat{S}_{KM}(t) = \prod_{\{j; \tau_j \leq t\}} (1 - 1/n_j) \text{ voor } t \in [0, T], \quad (1)$$

waarbij n_j het totaal aantal waarnemingen (gecensureerd of niet) groter dan of gelijk aan τ_j voorstelt. Voor zover er knopen zijn breken we die at random, maar plaatsen de gecensureerde waarnemingen in iedere knoop achter de niet gecensureerde. Voor T nemen we een of andere geschikt gekozen eindtijd, zodat $S^*(T) = S(T)G(T) > 0$, bijv. $T = \hat{S}^{*-1}(0.05)$, waarbij $\hat{S}^*$ de empirische overlevingsfunctie van alle waarnemingen is. Bij deze gegevens zouden we bijvoorbeeld ook 10 jaar als vast eindpunt kunnen nemen. Voor $t > T$ beschouwen we de Kaplan-Meier schatter als niet gedefinieerd. Een reden hiervoor is dat voor $t \rightarrow \sup_t \{t | G(t) > 0\}$ de asymptotische variantie van de KM schatter oneindig wordt. De asymptotische theorie is tamelijk ingewikkeld als we (bij toenemende steekproefgrootte) voor T steeds de grootste (of op k na grootste met k vast) waarneming kiezen (zie bijv. GILL (1983)). Voor de praktijk is dit laatste echter niet erg van belang, omdat we 'het venster' wel best iets kleiner kunnen maken. Een schatter voor de (asymptotische) variantie van $\hat{S}_{KM}(t)$ wordt gegeven door de volgende formule die, naar men zegt, voor het eerst door Greenwood (1926) is bedacht.

$$\hat{V}(\hat{S}_{KM}(t)) = \hat{S}_{KM}(t)^2 \sum_{\{j: \tau_j \leq t\}} (1/n_j(n_j-1)) \text{ voor } t \in [0, T]. \quad (2)$$

Een schatter analoog aan de Kaplan-Meier schatter, maar nu voor de cumulatieve hazard rate $\Lambda(t)$ is door Nelson (1968) geïntroduceerd:

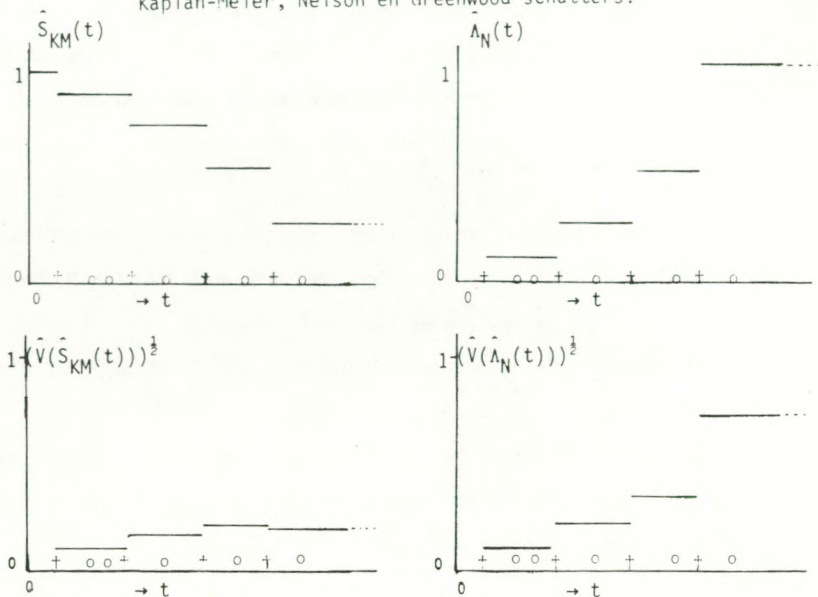
$$\hat{\Lambda}_N(t) = \sum_{\{j: \tau_j \leq t\}} (1/n_j) \text{ voor } t \in [0, T]. \quad (3)$$

De bijbehorende schatter voor de (asymptotische) variantie is

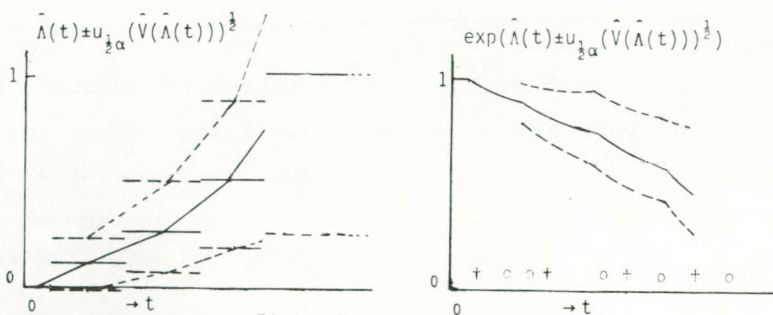
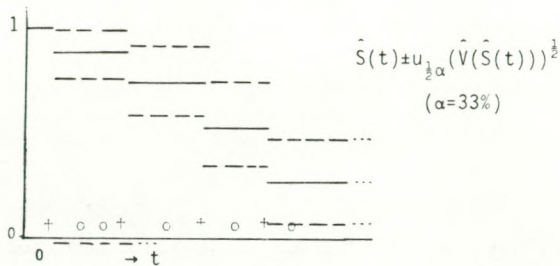
$$\hat{V}(\hat{\Lambda}_N(t)) = \sum_{\{j: \tau_j \leq t\}} (1/n_j(n_j-1)) \text{ voor } t \in [0, T]. \quad (4)$$

Er zij opgemerkt dat voor niet te kleine n_j geldt: $\hat{S}_N = \exp(\hat{\Lambda}_N) \approx \hat{S}_{KM}$. Merk ook het verband op tussen (2) en (4). Een recursieve versie van de formules (1) tot en met (4) is eenvoudig op een computer te programmeren. Figuur 1 geeft het uiterlijk van deze schatters in een concreet (hypothetisch) geval. Later in het verhaal zullen deze schatters vanuit een fundamenteeler gezichtspunt opnieuw opduiken.

Figuur 1.
Kaplan-Meier, Nelson en Greenwood schatters.



Figuur 2.
Instantane betrouwbaarheidsband voor de overlevingsfunctie.



Figuur 3.
Instantane betrouwbaarheidsband voor $\Lambda(t)$ en $S(t)$ ($\alpha = 33\%$).

Instantane betrouwbaarheidsbanden.

Na deze voorbereidingen is het gemakkelijk om (asymptotische) instantane betrouwbaarheidsbanden voor $S(t)$ te construeren. Een eerste mogelijkheid is $\hat{S}(t) \pm u_{1/2\alpha} (\hat{V}(\hat{S}(t)))^{1/2}$, t doorloopt $[0, T]$, te beschouwen. Dit levert enerzijds voor iedere vaste t een (asymptotisch) betrouwbaarheidsinterval voor $S(t)$ en anderzijds voor iedere vaste p een betrouwbaarheidsinterval voor het quantiel $S^{-1}(p)$. Dit laatste kunnen we inzien door te bedenken dat dit betrouwbaarheidsinterval beschouwd kan worden als het gebied (op de tijdas) waarvoor de nulhypothese $H_0: S(t)=p$ tegen $A: S(t) \neq p$ niet verworpen kan worden. Een dergelijk interval is direct uit de instantane betrouwbaarheidsband af te lezen (zie figuur 2). (In figuur 2 (en 3) is, om de betrouwbaarheidsband niet te breed te laten uitvallen, voor de asymptotische overschrijdingskans 33% genomen, wat overeenkomt met ' $\pm$ één keer de standaarddeviatie'). Op deze constructie kan voor de praktijk nog de volgende verfijning worden aangebracht: Vorm een betrouwbaarheidsband voor $\lambda(t) = -\ln S(t)$. Interpoleer de verkregen trapfunctie lineair en transformeer terug naar $S(t)$ (zie figuur 3). De lineaire interpolatie kan desgewenst ook op S worden toegepast. De asymptotische eigenschappen van de schatters veranderen natuurlijk niet door dergelijke interpolaties. De gedachte is dat de normale limiet verdeling een betere benadering voor de verdeling van $n^{1/2}(\hat{\Lambda}_N(t) - \Lambda(t))$ dan voor die van $n^{1/2}(\hat{S}_{KM}(t) - S(t))$ vormt (tenminste op het interval $[t_0, T]$ waarbij $\Lambda(t_0)$ niet te dicht bij 0 ligt). Simulatie experimenten van LINK (1979) schijnen hier ook op te wijzen. In THOMAS & GRUNKEMEIER (1975) worden nog 2 andere typen betrouwbaarheidsintervallen voorgesteld, die lastiger te berekenen zijn, maar bij simulatie experimenten betere eigenschappen vertonen. Niettemin zijn de volgende problemen (bij gecensureerde data) nog niet op bevredigende wijze opgelost.

(1) Voor welke 'eenvoudige' transformatie $\psi(S)$ wordt de exacte verdeling van $n^{1/2}(\hat{\psi}(S) - \psi(S))$ bij kleine steekproeven relatief goed door de normale limietverdeling benaderd?

(2) Welke bekende verdeling is, voor kleine steekproeven, een nauwkeurige benadering voor de exacte verdeling van $\hat{S}$? De beantwoording van deze vragen wordt vanzelfsprekend bemoeilijkt doordat een en ander af kan hangen van de censureringsverdeling G.

Simultane betrouwbaarheidsbanden (discreet).

Beschouw een vaste partitie $0 < s_1 < s_2, \dots, s_k < T$ van de tijdas ($k < n$). Bijvoorbeeld zou men kunnen kijken naar de 1, 3, 5 en 8 jaars overlevingspercentages. We gaan gebruik maken van asymptotische theorie, in de afgelopen 10 jaar door verschillende mensen ontwikkeld, waarbij $\hat{S}(t)$ en $\hat{\Lambda}(t)$ als stochastische processen (toevalsafhankelijke tijdsfuncties) worden opgevat. Onder algemene voorwaarden geldt dat $Y_n(t) = n^{1/2}(\hat{S}(t) - S(t)) \xrightarrow{D} Y(t)$, waarbij $Y(t)$ een gaussisch proces is met $EY(t) = 0$ en covariantiestructuur $EY(s)Y(t) = S(s)S(t)A(\min(s, t))$. Hierbij kan de censureringsintegraal $A(t) = -\int_0^t (S^2(u)G(u))^{-1} dS(u)$ consistent geschat worden door $\sum_{j: \tau_j \leq t} (1/n_j(n_j-1))$. Deze schatter kan verkregen worden door voor $S(u)$ de KM schatter en voor $S(u)G(u)$ de empirische overlevingsfunctie van alle waarnemingen (gecensureerd of niet) in te vullen. De convergentiebegrip $\xrightarrow{D}$ houdt ondermeer in dat alle eindig dimensionale verdelingen van $Y_n(t)$ naar de corresponderende marginale verdeling van $Y(t)$ convergeren. Dus $(Y_n(s_1), \dots, Y_n(s_k))$ nadert in verdeling naar $N(0, \Sigma)$, waarbij $\Sigma_{ij} = A(\min(s_i, s_j))$ te schatten is door de Greenwood formule voor Λ . De multivariaat normale theorie is vertrouwd terrein, zie bijv. JOHNSON & KOTZ (1978) en MILLER (1981b). Zoals bekend is het moeilijk de overschrijdingskansen $p(c_1, \dots, c_k) = 1 - P\{|Y(s_j)| < c_j \text{ voor alle } j=1, \dots, k\}$ exact te bepalen. Ze zijn echter wel (conservatief) af te schatten door de Bonferroni ongelijkheid: $p(c_1, \dots, c_k) < \sum_{j=1}^k \alpha_j$ met $\alpha_j = P\{|Y(s_j)| > c_j\}$. We zien nu dat een discrete simultane betrouwbaarheidsband overeenkomt met een aanpassingstoets met een speciale vorm van het verwerpsgebied. De nulhypothese $S(s_j) = S_0(s_j)$ voor alle $j=1, \dots, k$ wordt verworpen als $S_0(s_j)$ voor tenminste één s_j buiten het betrouwbaarheidsinterval $\hat{S}(s_j) \pm n^{-1/2}c_j$ valt. Dit leidt

voor $n^{1/2}(S_0(s_1) - \hat{S}(s_1)), \dots, n^{1/2}(S_0(s_k) - S(s_k))$ in R^k tot een balkvormig acceptatiegebied. Deze vorm van het acceptatiegebied, veroorzaakt doordat men de toets in een tweedimensionale grafiek direct wil kunnen uitvoeren, leidt tot een moeilijke verdeling van de toetsgrootte. Beschouwen we het goodness-of-fit probleem zonder deze 'balkbeperking', dan zijn er asymptotisch natuurlijk allerlei bekende toetsen uit te voeren. Bijvoorbeeld de χ^2 toets tegen het niet ingeperkte alternatief $A_1: S(s_i) \neq S_0(s_i)$, en most stringent lineaire of likelihood ratio toetsen tegen het ingeperkte alternatief $A_2: S(s_i) < S_0(s_i)$ met tenminste één ongelijkheid strict (dit is een Groningse interesse). Een opmerking in dit verband is nog dat, wanneer men $Z_n(t) = n^{1/2}(\hat{\Lambda}(t) - \Lambda(t))$ beschouwt in plaats van $Y_n(t)$, het limietproces $Z(t)$ een gaussisch proces met onafhankelijke aangroeiingen is. Dit wordt ook wel een (tijdsgetransformeerde) Brownse beweging genoemd. Het houdt in dat de covariantiematrix van $(Z(s_1), \dots, Z(s_k))$ op eenvoudige manier (nl. door over te gaan op de verschillen: $Z(s_1), Z(s_2) - Z(s_1), \dots, Z(s_k) - Z(s_{k-1})$) gediagonaliseerd kan worden. Toetsen voor $Y(t)$ kunnen worden omgevormd tot toetsen voor $Z(t)$, waardoor sommige formules eenvoudiger worden.

Simultane betrouwbaarheidsbanden (continu).

Over dit onderwerp zal ik kort zijn. Een basisgedachte, die in verschillende variaties is uitgewerkt, is de volgende: $n^{1/2}(\hat{S}(t) - S(t))/S(t)$ convergeert, evenals $n^{1/2}(\hat{\Lambda}(t) - \Lambda(t))$, naar een tijdsgetransformeerde Brownse beweging.

Voor de Brownse beweging zijn de overschrijdingskansen

$P\{W(t) \in (c_1 + d_1 t, c_2 + d_2 t) \quad \forall t \in [0, T]\}$ uit te drukken in een reeksontwikkeling (ANDERSON, (1960)). Invullen van een schattter $\hat{A}(t)$ voor $A(t)$ en terugtransformeren naar $\hat{S}(t) - S(t)$ levert een (familie van) betrouwbaarheidsbanden. GILLESPIE & FISHER (1979) hebben hebben dit soort banden voor $S(t)$ bij gecensureerde data voorgesteld. De berekening van overschrijdingskansen voor algemenere grenzen is nog voorwerp van onderzoek. Resultaten zijn o.a. te vinden in JENNEN (1982) en JENNEN & LERCHE (1981),

zie ook SCHUMACHER (1982a). Een variatie op dit thema is, dat $n^{1/2}(\hat{S}(t)-S(t))/S(t)(1+A(t))$ naar een tijdsgetransformeerde Brownse brug (gaussisch proces met verwachting nul en covariantiestructuur $K(\min(s,t))-K(s)K(t)$, waarbij $K(t)=A(t)/(1+A(t))$) convergeert. De overschrijdingskansen van een Brownse brug zijn te herleiden tot die van een Brownse beweging. Het corresponderende type betrouwbaarheidsband is door HALL & WELLNER (1980) voorgesteld. BUNDSCHUH (1982) heeft een interessante simulatiestudie naar beide typen banden verricht, waaruit blijkt, dat de 'Hall en Wellner' banden in een aantal gevallen betere kleine steekproef eigenschappen hebben dan de 'Gillespie en Fisher' banden. Nog niet naar bevrediging opgeloste problemen zijn:

(1) Het vinden van een geschikte vorm voor de betrouwbaarheidsbanden. Enerzijds moet een dergelijke vorm relevant zijn voor de praktische toepassingen en anderzijds moeten de overschrijdingskansen berekenbaar blijven. Bij het bepalen van de vorm zou ook mee moeten spelen tegen welke alternatieven men bij de corresponderende aanpassingstoets een hoog onderscheidingsvermogen wil hebben.

(2) Het bepalen voor welke steekproefgrootten de asymptotische theorie een adequate benadering geeft.

Simulatie experimenten kunnen op vraag 2 gedeeltelijk een antwoord geven. Op het iets ruimere probleem van goodness-of-fit in deze context wil ik hier niet verder in gaan. Werk aan verschillende typen toetsgrootheden is hierover in volle gang (zie o.a. HYDE (1977), HOLLANDER & PROSCHAN (1979), KOZIOL & GREEN (1976), NAIR (1981), PETTIT & STEPHENS (1976), SCHUMACHER (1982b)).

Toepassing.

Tot slot zullen we een gedeelte van de theorie op de concrete gegevens toepassen. Wanneer we van de in de inleiding genoemde levensduren de Kaplan-Meier schatter uitrekenen (met de Greenwood schatter voor de variantie) en deze in een grafiek uitzetten samen met de overlevingscurven van de

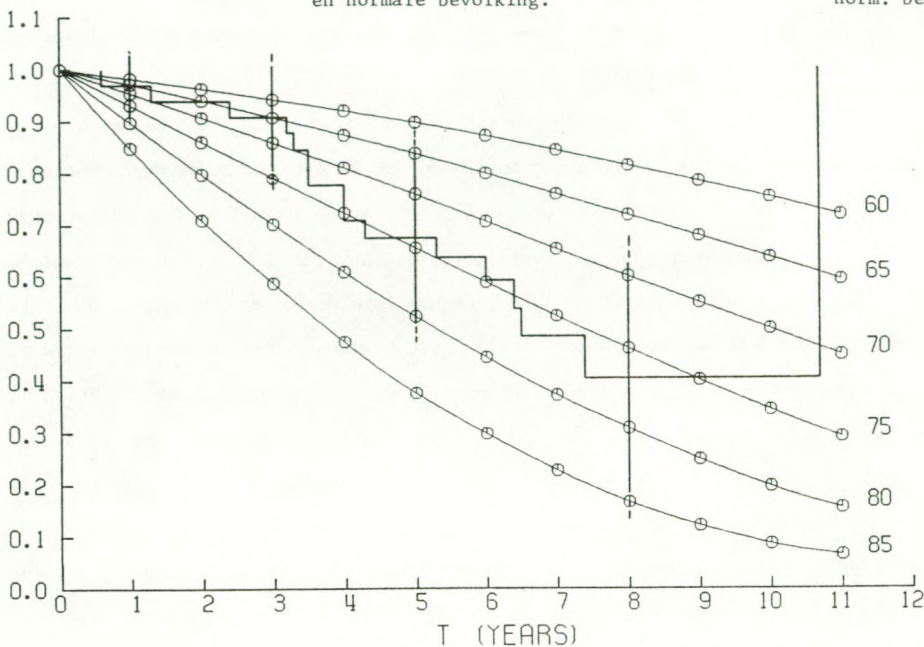
normale Nederlandse mannen (C.B.S. (1976)), dan krijgen we een tekening als in figuur 4. De verticale lijnen (men zou er wat meer moeten tekenen) kunnen gebruikt worden als instantane betrouwbaarheidsband, met de eerder vermelde interpretatie. Wanneer we een (discrete) simultane band willen beschouwen (in de tekening voor 4 punten), dan moeten de verticale foutengrenzen verlengd worden. Met de Bonferroni ongelijkheid (voor ieder van de 4 tijdstippen $s_1, \dots, s_4$ (asymptotisch) een overschrijdingskans van $5/4\%$ en symmetrische staarten) levert dit in de figuur de gestippelde uitbreidingen. De discrete simultane betrouwbaarheidsbanden zijn te gebruiken als goodness-of-fit toets. Bijvoorbeeld $H_0: S(t) = S_0(t|80)$ (de normale manlijke bevolking met leeftijd 80 jaar) wordt in dit geval verworpen, omdat de grafiek $S_0(t|80)$ op tenminste één plaats (bij s_2) buiten de gestippelde intervallen valt. Indien we de overlevingscurven van de normale bevolking, geparametriseerd door de leeftijd, als 'referentiecurven' willen aanvaarden, dan komen we (op grond van deze speciale band gebaseerd op 4 tijdstippen) tot het volgende 95% betrouwbaarheidsinterval van de niet verworpen nulhypothese: $(S_0(t|67), S_0(t|77))$.

In figuur 5 zijn rond de Kaplan-Meier schatter betrouwbaarheidsbanden volgens de methode van Hall en Wellner getekend (asymptotisch betrouwbaarheidsniveau 95%), waarbij (lineair) is geïnterpoleerd door de middens van de optredende horizontale lijnstukken met elkaar te verbinden. Ook deze banden kunnen als aanpassingstoets voor $H_0: S(t) = S_0(t)$, $0 < t < T$, worden gebruikt en leveren als gebied van de niet verworpen nulhypothese: $(S_0(t|65), S_0(t|82))$. Het uitrekenen van deze banden vergt meer werk dan de eerste, snellere en ruwere, methode. De resultaten stemmen in redelijke mate overeen. Verschillen zijn toe te schrijven aan: (1) het feit dat bij de eerste methode slechts naar 4 (van te voren gekozen) tijdstippen wordt gekeken; (2) het feit dat de overlevingscurven van normale Nederlanders geen goede referentiecurven hoeven te zijn voor die van de patienten; (3) de toepassing van asymptotische methoden bij een tamelijk geringe steekproefomvang. Er

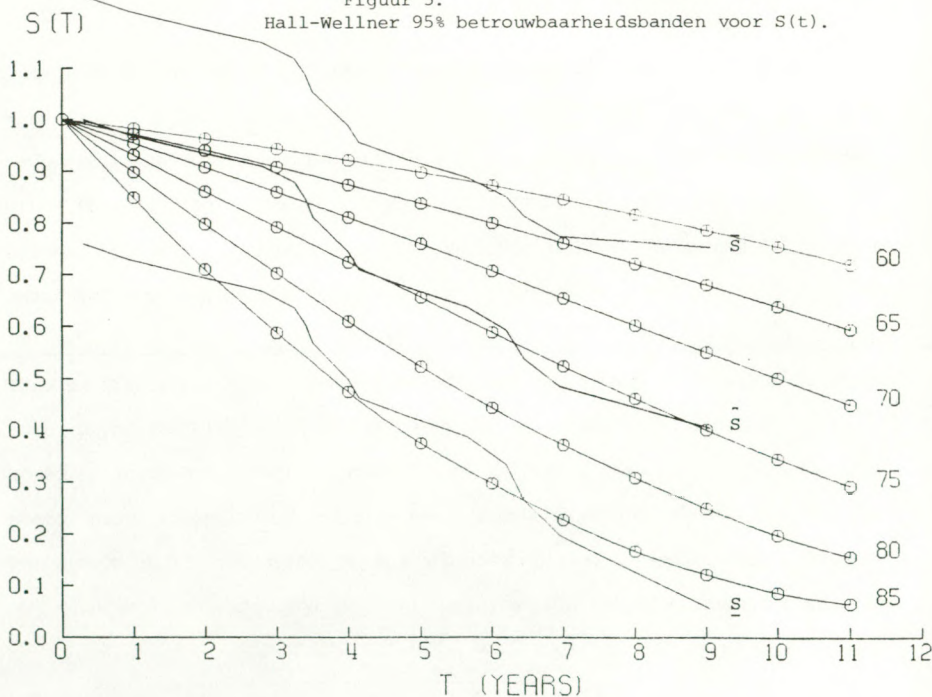
14
S(T)

Figuur 4.
Overleving stadium I larynx-kanker patienten
en normale bevolking.

leeftijd
norm. bev.



Figuur 5.
Hall-Wellner 95% betrouwbaarheidsbanden voor $S(t)$.



valt op dat de lage waarde van de bovenste leeftijdsgrens bij de eerste band samenhangt met de (toevallig?) hoge waarde van de geschatte overlevingsfunctie in de buurt van 3 jaar.

Uit de voorgaande berekeningen blijkt dat volgens deze waarnemingen de patiënten met stadium I larynx-kanker, ongeacht hun eigen leeftijd, ruwweg dezelfde overlevingskansen hebben als de normale manlijke bevolking met een vaste leeftijd, die wij niet precies kennen, maar die tussen de 65 en 82 jaar zal liggen. Dit geldt voor de eerste 10 jaar en ongeveer voor de leeftijdsklasse, waarin de waargenomen patiënten zich bevinden jaar). Het interval (65,82) zou men dus kunnen opvatten als een betrouwbaarheidsinterval voor de 'effectieve leeftijd' van de patiënten ten gevolge van de aanwezigheid van de kanker. Uiteraard kunnen geen al te stringente conclusies uit dit interval getrokken worden. Niettemin lijkt het erop dat, simplificerend gesproken, de oudere patiënten (> 65 jaar) geen noemenswaardige levensduurverkorting ten gevolge van de kanker ondergaan, terwijl dit met de jongere patiënten wel het geval is. (Van de 33 patiënten zijn er 16 jonger dan 65 en is er 1 ouder dan 82). Verder zij er nog opgemerkt dat het onderzoek retrospectief is en dat de behandelingswijze niet als verklarende variabele is opgenomen, zodat het interessant blijft deze conclusie aan andere, onafhankelijke, onderzoeken te toetsen.

De aanname dat de overlevingscurve van de patiënten in de eerste 10 jaar onafhankelijk is van hun eigen leeftijd, is een vereenvoudiging die we, met in de inleiding genoemde argumentatie, hebben aangebracht. Wanneer deze vereenvoudiging niet mag worden gemaakt, dan worden de zaken aanmerkelijk ingewikkelder.

Het hoofdmotief van de praktische toepassing was dus niet om uit te maken of de patiënten van stadium I over het algemeen sneller sterven dan normale Nederlanders, door de patiënten ieder met een normale leeftijdsgenoot te vergelijken (dit staat elders uitgewerkt, zie bijv. KARDAUN (1983)), maar om een betrouwbaarheidsinterval voor de 'effectieve leeftijd' van de patiënten te

vinden, in de benadering dat hun eigen leeftijd er wat dit betreft niet zoveel toe doet.

Ik ben de deelnemers van de MBS workshop in Heeze, prof. dr. W. Schaafsma en prof. dr. A.J. Stam erkentelijk voor hun opbouwende kritiek en bijdrage tot de discussie. Ik dank dr. J.H. Pasmooij en dr. H.C. Stam voor toestemming tot het publiceren van de data en de vakgroep Informatica voor het gebruik van de diablo.

Literatuur.

- ANDERSEN, P.K., BORGAN, Ø., GILL R.D., & KEIDING N. (1982). Linear non-parametric tests for comparison of counting processes, with applications to censored survival data (with discussion), *International Statistical Review* 50, 219-259.
- ANDERSON, T.W. (1960). A modification of the sequential probability ratio test to reduce the sample size, *Annals of Mathematical Statistics* 31, 165-197.
- BARR, D.R. & DAVIDSON, T. (1973). A Kolmogorov-Smirnov test for censored samples, *Technometrics* 15, 739-757.
- BRESLOW, N.E. & CROWLEY, J. (1974). A large sample study of the life table and product limit estimates under random censorship, *Annals of Statistics* 2, 437-453.
- BROOKMEYER, R. & CROWLEY, J. (1982). A confidence interval for the median survival time, *Biometrics* 38, 29-41.
- BUNDSCHUH, W. (1982). Simultane Konfidenzbänder für die Survivalfunktion, Diplomarbeit, Fakultät für Mathematik, Heidelberg.
- BURKE, M.D., CSÖRÖG, S. & HORVÁTH, L. (1982). Strong approximations of some biometric estimates under random censorship, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 56, 87-112
- Centraal Bureau voor de Statistiek (1976). Lifetable Dutch males, Staatsuitgeverij, 's-Gravenhage.
- CHEN, Y.Y., HOLLANDER, M. & LANGBERG, N.A. (1982). Small sample results for the Kaplan-Meier estimator, *Journal of the American Statistical Association* 77, 141-144.
- EMERSON, J.D. (1982), Nonparametric confidence intervals for the median in the presence of right censoring, *Biometrics* 38, 17-27.
- FLEMING, T.R., O'FALLON, J.R., O'BRIEN, P.C. & HARRINGTON, D.P. (1980). Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right censored data, *Biometrics* 36, 607-625.
- FLEMING, T.R. & HARRINGTON, D.P. (1981). A class of hypothesis tests for one and two sample censored survival data, *Communications in Statistics A* 10, 763-794.
- FÖLDES, A. & RETJÖ, L. (1981). Strong uniform consistency for nonparametric survival curve estimators from randomly censored data, *Annals of Statistics* 9, 122-129.
- GILLESPIE, M.J. & FISHER, L. (1979). Confidence bands for the Kaplan-Meier survival curve estimate, *Annals of Statistics* 7, 920-924.
- GILL, R.D. (1980). Censoring and stochastic integrals, *M.C. Tracts* 124, Mathematisch Centrum, Amsterdam.
- GILL, R.D. (1983). Large sample behaviour of the product-limit estimator on the whole line, *Annals of Statistics* 11, 49-59.

- GREENWOOD, M. (1926). The natural duration of cancer, reports on public health and medical subjects, 33, London, Her Majesty's Stationary Office, 1-26.
- HALL, W.J. & WELLNER, J.A. (1980). Confidence bands for a survival curve from censored data, *Biometrika* 67, 133-143.
- HYDE, J. (1977). Testing survival under right censoring and left truncation, *Biometrika* 64, 225-230.
- HOLLANDER, M & PROSCHAN, F. (1979). Testing to determine the underlying distribution using randomly censored data, *Biometrics* 35, 393-401.
- JENNEN, C. & LERCHE, H.R. (1981). First exit densities of brownian motion through one-sided moving boundaries, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 55, 133-148.
- JENNEN, C. (1982). Asymptotische bestimmung von kengrößen sequentieller Verfahren, Dissertation, Universität Heidelberg.
- JOHANSEN, S. (1978). The product-limit estimator as maximum likelihood estimator, *Scandinavian Journal of Statistics* 5, 195-199.
- JOHNSON, N.L. & KOTZ, S. (1972). Distributions in statistics: continous multivariate distributions, J. Wiley, New-York, London.
- KALBFLEISCH, J.D. & PRENTICE, R.L. (1980). The statistical analysis of failure time data, J. Wiley, New-York, Toronto.
- KAPLAN, E.L. & MEIER, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association* 53, 457-481.
- KARDAUN, O.J.W.F. (1983). Statistical survival analysis of male larynx-cancer patients, a case study, *Statistica Neerlandica* 37, 3, 103-125.
- KOZIOL, J.A. & BYAR, D.P. (1975). Percentage points of the asymptotic distributions of one and two sample K-S statistics for truncated or censored data, *Technometrics* 17, 507-510.
- KOZIOL, J.A. & GREEN, S.B. (1976). A Cramèr-von Mises statistic for randomly censored data, *Biometrika* 63, 465-474.
- LINK, C.L. (1979). Confidence intervals for the survival function using Cox's proportional hazards model with covariates, Technical report no. 45 (R01 GM21215), Division of Biostatistics, Stanford University, Stanford, California.
- MILLER, R.G. Jr. (1981a). Survival Analysis, J. Wiley, New-York.
- MILLER, R.G. Jr. (1981b). Simultaneous statistical inference (2nd edition), Springer Verlag, New-York, Heidelberg.
- NAIR, V.N. (1981). Plots and tests for goodness of fit with randomly censored data, *Biometrika* 68, 99-103.
- NELSON, W. (1968). Hazard plotting for incomplete failure data, *Journal of Quality Technology* 1, 27-52.
- NELSON, W. (1972). Theory and application of hazard plotting for censored failure data, *Technometrics* 14, 945-966.

- PETTIT, A.N. & STEPHENS, M.A. (1976). Modified Cramèr-von Mises statistics for censored data, *Biometrika* 63, 291-298.
- REID, N. (1981a). Influence functions for censored data, *Annals of Statistics* 9, 78-92.
- REID, N. (1981b). Estimating the median survival time, *Biometrika* 68, 188-190.
- SCHUMACHER, M. (1982a). Analyse von "Überlebenszeiten bei nichtproportionalen Hazardfunktionen, Habilitationsschrift, Fakultät für Theoretische Medizin, Heidelberg.
- SCHUMACHER, M. (1982b). Two-sample tests of Cramèr-von Mises and Kolmogorov-Smirnov type for randomly censored data, preprint 195, Sonderforschungsbereich 123, Heidelberg.
- THOMAS, D.R. & GRUNKEMEIER, G.L. (1975). Confidence interval estimation of survival probabilities for censored data, *Journal of the American Statistical Association* 70, 865-871.
- TURNBULL, B.W. & WEISS, L. (1978). A likelihood ratio statistic for testing goodness of fit with randomly censored data, *Biometrics* 34, 367-375.

Ontvangen: 29-8-1983