stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SN 6/77 AUGUSTUS

J.G. BETHLEHEM, R.J.M.M. DOES & R.D. GILL

VERDELINGSVRIJE METHODEN BIJ CENSURERING

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

INHOUD

VOORWOORD

Dit rapport heeft als syllabus gediend bij de werkweek "Stochastische Censurering" georganiseerd door de afdeling Mathematische Statistiek in augustus 1977.

In hoofdstuk 1 worden drie verdelingsvrije toetsen voor het twee-steekproeven probleem in geval van censurering beschreven. Bovendien wordt in die situatie een methode gegeven om de verdelingsfunctie verdelingsvrij te schatten.

In hoofdstuk 2 wordt de theorie der stochastische integralen en telprocessen ontwikkeld. Voor een uitgebreide behandeling van de stochastische integralen zie MEYER [20].

In het laatste hoofdstuk wordt de theorie van telprocessen toegepast om eigenschappen van de toetsen uit het eerste hoofdstuk te bewijzen en om een asymptotische vergelijking van deze toetsen te maken. De techniek om deze toetsen in termen van stochastische integralen en telprocessen te schrijven, is afkomstig van AALEN [1] en [2].

In dit rapport is voor het eerst een rigoreus bewijs geleverd voor de asymptotische normaliteit van de voorwaardelijke toets van Gehan en voor het konsistent zijn van deze toets. De asymptotische resultaten uit hoofdstuk 3 overlappen de bekende resultaten. Wij eisen voor de verdelingsfunctie een dichtheid, maar van de verdeling van de censureringsvariabelen eisen wij vrijwel niets. Bovendien is type II censurering toegelaten.

Wij zijn veel dank verschuldigd aan prof.dr. J. Oosterhoff voor zijn adviezen bij het schrijven van dit rapport, evenals aan drs. P. Groeneboom.

Amsterdam, augustus 1977.

J.G. Bethlehem,
R.J.M.M. Does,
R.D. Gill.

HOOFDSTUK 1

1.1. INLEIDING

Levensduurtabellen vinden al zeer lang toepassing als statistische techniek. Dit gebeurt niet alleen in medisch onderzoek, maar ook in de industrie is men vaak geïnteresseerd in de levensduur van objecten (mensen, planten, gloeilampen, etc.). Wat men eigenlijk wil weten is de levensduurverdeling van de betreffende objecten, dus, in statistische terminologie, de verdeling van de stochastische variabele X die de levensduur voorstelt.

DEFINITIE 1.1.1. De *overlevingsfunctie* wordt gedefinieerd door:

$$S(x) = P(X>x) = 1 - F(x)$$

waarin F de verdelingsfunctie van de stochast X is. In woorden: S(x) is de kans dat een object een levensduur groter dan x heeft.

In het algemeen zullen we niet over levensduren praten, maar over de tijd die verstrijkt tot een bepaalde specifieke gebeurtenis optreedt.

Men kan zich voorstellen hoe men in de praktijk te werk moet gaan om S voor een bepaalde populatie te schatten: Trek een aselecte steekproef en bepaal van ieder element de tijd tot het optreden van de specifieke gebeurtenis. Dit levert een aantal waarden van $X_1,\dots,X_n$. Schat S(x) dan met

$$S_n(x) = \frac{\#\{i:X_i>x\}}{n} .$$

In de praktijk treden echter vaak complicaties op die het schatten van S bemoeilijken. Een veel voorkomend probleem is dat men niet van ieder object precies de verstreken tijd tot het optreden van de specifieke gebeurtenis kan bepalen, omdat het optreden van een andere gebeurtenis dit onmogelijk maakt.

Wanneer een dergelijke situatie optreedt, spreken we van *censurering*.

We kunnen censurering als volgt formeel in een model beschrijven: Noem de verstreken tijd tot het optreden van de specifieke gebeurtenis de wachttijd. Uit een populatie is een aselecte steekproef van omvang n getrokken. De wachttijden van de objecten in de steekproef geven we aan met $X_1, X_2, \ldots, X_n$. Bovendien zijn er n stochasten $U_1, U_2, \ldots, U_n$, die voor het corresponderende object het tijdstip van censurering bepalen. Men neemt nu het volgende waar:

(a) $\tilde{X}_i \overset{\text{def}}{=} X_i \wedge U_i \overset{\text{def}}{=} \min\{X_i, U_i\}, \qquad i = 1,2,\ldots,n$

(b) $\Delta_i \overset{\text{def}}{=} I_{\{X_i \leq U_i\}}, \qquad i = 1,2,\ldots,n$ (I is de indicatorfunctie).

D.w.z. men neemt waar de tijd die verstrijkt tot het optreden van de specifieke gebeurtenis òf een andere gebeurtenis die verdere waarneming onmogelijk maakt. Bovendien weet men of het de specifieke gebeurtenis was of een andere gebeurtenis, die de waarneming beëindigde.

Al naar gelang het karakter van de censureringsstochasten $U_1, U_2, \ldots, U_n$ kunnen we verschillende gevallen van censurering aangeven:

(i) Vaste waarnemingsgrens: $U_i \equiv u$, $i = 1,2,\ldots,n$. Alle waarnemingen worden, voor zover de specifieke gebeurtenis nog niet is opgetreden, op een vast tijdstip beëindigd.

(ii) Vaste censurering: $U_i \equiv u_i$, $i = 1,2,\ldots,n$. Voor ieder object is een vaste maar verschillende waarnemingsgrens vastgelegd.

(iii) Identieke censurering: De U_i's zijn onderling onafhankelijk (o.o.) en identiek verdeeld. Bovendien zijn ze onafhankelijk van de X_i's.

(iv) Ongelijke censurering: De U_i's zijn o.o. maar niet noodzakelijk identiek verdeeld en ze zijn onafhankelijk van de X_i's.

(v) Willekeurige censurering: De U_i's zijn afhankelijk, van elkaar en van de X_i's en niet noodzakelijk identiek verdeeld.

In gevallen (iii), (iv) en (v) spreken we van *stochastische censurering*. De U_i's worden *concurrerende risico's* genoemd.

Geval (i) noemt men in de literatuur wel *censurering van type I*. Men spreekt dan van *censurering van type II* wanneer geldt: $U_i = X_{(r)}$, $i = 1,2,\ldots,n$;

$1 \leq r \leq n$, waarin $X_{(1)},\ldots,X_{(n)}$ de geordende steekproef is. Dit valt onder geval (v).

Een andere gebruikte terminologie is de volgende:
Men spreekt van Model I-censurering ("random censorship"), wanneer in een k-steekproevenprobleem de verdeling van de censureringsstochasten mag verschillen van steekproef tot steekproef, maar binnen een steekproef identiek is. Dit is per steekproef geval (iii).
Model II-censurering ("fixed censorship") is in feite geval (ii). Indien de U_i's een willekeurige simultane verdeling hebben die onafhankelijk is van de X_i's, dan verkeert men, voorwaardelijk op de gerealiseerde waarden van de U_i's, ook in geval (ii)

We willen hier, misschien ten overvloedde, nog opmerken dat censurering niet hetzelfde is als afknotting (afkapping). Afknotting is een eigenschap van de onderliggende verdeling: We doen alleen waarneming afkomstig uit een beperkt gebied; We trekken als het ware een steekproef uit een voorwaardelijke verdeling. Censurering daarentegen is een eigenschap van de steekproef: we kennen alleen waarnemingen uit een beperkt gebied en weten bovendien dat er (en hoeveel) waarnemingen buiten dit gebied liggen.

We illustreren censurering nu eerst aan de hand van enkele voorbeelden.

<u>VOORBEELD 1.1.2.</u> Ratten
Men wil iets weten over de tijd die verstrijkt tussen het inenten van ratten met een bepaalde stof en het verschijnen van een tumor als gevolg daarvan. Van een aantal proefdieren zal men deze tijdsduur zonder meer kunnen meten, maar het is goed denkbaar dat dit bij bepaalde dieren niet mogelijk is. Oorzaak: Ze sterven voor die tijd aan een andere doodsoorzaak. Ze ontvluchten bijvoorbeeld uit hun hok.

<u>VOORBEELD 1.1.3.</u> Harttransplantaties
Men wil van patiënten, die een harttransplantatie hebben ondergaan, nagaan hoe lang ze nog te leven hebben. Oorzaken die tot censurering kunnen leiden, zijn: Sterven aan een andere doodsoorzaak, verdwijnen uit administratie als gevolg van ontslag uit het ziekenhuis, afsluiten van het onderzoek na een bepaalde periode als gevolg van geldgebrek, etc.

<u>VOORBEELD 1.1.4.</u> Signaal
Men wil de grootte van een signaal meten dat door een bepaald electronisch apparaat wordt afgegeven. Een vervelende complicatie daarbij is de aanwezigheid van ruis, zodat beneden een bepaalde waarde het signaal onmeetbaar wordt

door de overheersende ruis. Hierbij treedt op wat we in feite linker-censurering moeten noemen. Wat we waarnemen is $\max\{X_i,U_i\}$, waarin X_i het signaal en U_i de ruis. We weten dus dat de wachttijd hoogstens gelijk is aan $\max\{X_i,U_i\}$; Bij de door ons al eerder gedefinieerde censurering nemen we $\min\{X_i,U_i\}$ waar en weten we dat de wachttijd minstens gelijk is aan $\tilde{X}_i$. We zouden dit eigenlijk rechter-censurering moeten noemen. De twee hier beschreven situaties zijn volledig analoog. Daarom zullen we alleen over rechter-censurering spreken en daarbij het voorvoegsel rechter-weglaten.

We zullen ons in het bijzonder gaan bezighouden met het verdelingsvrije twee-steekproevenprobleem. We hebben twee aselecte steekproeven, getrokken onder verschillende omstandigheden. Onze nulhypothese H_0 luidt dat de verdeling van de wachttijden in beide situaties dezelfde is. De alternatieve hypothese kan één- of tweezijdig zijn: de kans dat een willekeurig element uit de ene steekproef een grotere wachttijd heeft dan een willekeurig element uit de andere steekproef is groter dan/kleiner dan/ongelijk aan $\frac{1}{2}$.

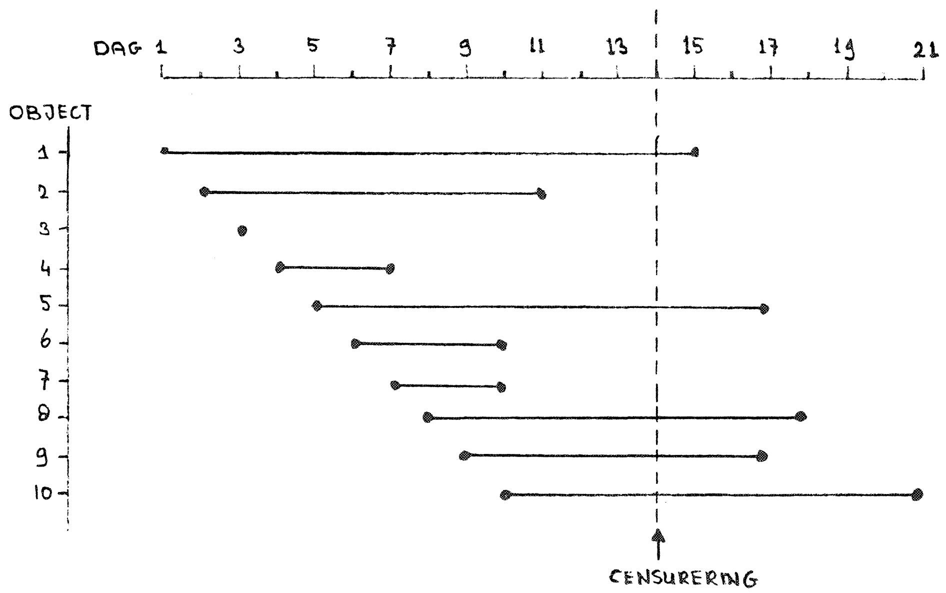
Dus, als we teruggaan naar de voorbeelden, willen we de volgende vragen beantwoorden:
"Zal bij het gebruik van stof A een tumor eerder optreden dan bij gebruik van stof B?", "Zal door een extra behandeling de patiënt met een nieuw hart langer leven dan zonder die behandeling?" en "Geeft het ene apparaat een groter signaal af dan het andere apparaat?"

We zullen drie belangrijke artikelen over het twee-steekproevenprobleem, die van GEHAN [14], EFRON [13] en COX [7], behandelen. Daarna zullen we in een wat meer theoretisch gedeelte dieper ingaan op de theorie van "telprocessen", zoals die is ontwikkeld door AALEN [1]. Deze theorie levert ons het gereedschap, waarmee een goede vergelijking van verschillende methoden mogelijk is. Tenslotte zullen we die vergelijking ook inderdaad maken.

Een eenvoudige, voor de hand liggende manier om het probleem van gecensureerde steekproeven op te lossen, lijkt het buiten beschouwing laten van de gecensureerde waarnemingen. In het algemeen is dit een niet aan te bevelen procedure, zoals uit onderstaand voorbeeld zal blijken. In die gevallen, waarin het wel wordt toegepast, levert het, door het geringere aantal waarnemingen, veel minder nauwkeurige schatters.

<u>VOORBEELD 1.1.5.</u> Yoghurt
We willen aan de hand van een steekproef ter grootte 10 de kans schatten dat een fles yoghurt minstens 5 dagen goed blijft (notatie: p_5). Om de

melkboer niet te overbelasten kiezen we op 10 achtereenvolgende dagen aselect een fles bij hem uit. De gekozen flessen yoghurt blijken de volgende levensduur te hebben (In feite zijn de getallen aselect getrokken uit de gehele getallen 0 t/m 15, zodat de theoretische kans gelijk is aan $p_5 = \frac{11}{16} = 0{,}69$): 14, 9, 0, 3, 12, 4, 3, 10, 8, 11.
In een grafiek krijgen we het volgende patroon:



We schatten p_5 met $\hat{p}_5 = \frac{\#\{i : X_i \geq 5\}}{n}$ en vinden als waarde 0,60. Het 90% betrouwbaarheidsinterval wordt gegeven door: (0,30, 0,85). Stel nu echter eens dat we het experiment op de 14-de dag hadden moeten beëindigen, gedwongen door omstandigheden. Dan hadden we de volgende waarnemingen gehad (gecensureerde waarnemingen worden aangegeven door *): 13*, 9, 0, 3, 9*, 4, 3, 6*, 5*, 4*. Als we nu de gecensureerde waarnemingen weggooien, dan houden we over: 9, 0, 3, 4, 3.

We schatten p_5 nu met 0,20 en het betrouwbaarheidsinterval is nu (0,01, 0,66). De schatting is volledig fout. De oorzaak is duidelijk. De gecensureerde waarnemingen, die we weg hebben gegooid, waren juist de grote waarnemingen. Het is dus logisch dat we systematisch te laag schatten. Zeker in deze situatie is het ook zonde om de gecensureerde waarnemingen weg te gooien, want van 4 van de 5 gecensureerde objecten weten we dat hun levensduur minstens 5 dagen was, zodat we ze in onze schatting kunnen betrekken. Als we alleen de laatste waarneming weggooien, dan kunnen we p_5 schatten met 5/9 = 0,56. Het betrouwbaarheidsinterval is (0,25, 0,83).

Alvorens nu tot behandeling van een aantal verdelingsvrije twee-steekproevenmethoden over te gaan, voeren we eerst in de volgende tabel een aantal notaties in.

Notaties

OMSCHRIJVING	EERSTE STEEKPROEF	TWEEDE STEEKPROEF
Aantal waarnemingen:	m	n
Wachttijden:	$X_1,X_2,...,X_m$	$Y_1,Y_2,...,Y_n$
Verdelingfunctie (indien wachttijden identiek verdeeld):	F	G
Bijbehorende overlevingsfunctie:	S	T
Censurerende stochasten:	$U_1,U_2,...,U_m$	$V_1,V_2,...,V_n$
Verdelingsfunctie (indien censurerende stochasten identiek verdeeld):	J	K
Waarnemingen:	$\tilde{X}_1,\tilde{X}_2,...,\tilde{X}_m$	$\tilde{Y}_1,\tilde{Y}_2,...,\tilde{Y}_n$
Verdelingsfunctie van $\tilde{X}_i$'s resp. $\tilde{Y}_j$'s (indien identiek verdeeld):	$\tilde{F}$	$\tilde{G}$
Censuur-indicator (0, indien gecensureerd):	$\Delta_1,\Delta_2,...,\Delta_m$	$E_1,E_2,...,E_n$
Subverdelingsfunctie van ongecensureerde waarnemingen ($P(\tilde{X}_i\leq x,\Delta_i=1)$):	F^0	G^0
Empirische schatting voor F resp. G:	F_m	G_n
Empirische schatting voor $\tilde{F}$ resp. $\tilde{G}$:	$\tilde{F}_m$	$\tilde{G}_n$
Empirische schatting voor F^0 resp. G^0:	F^0_m	G^0_n

1.2. DE TOETS VAN GEHAN

GEHAN [14] heeft een toets geconstrueerd naar analogie van de twee-steekproeven toets van Wilcoxon. D.w.z. hij vergelijkt paren $(\tilde{X}_i,\tilde{Y}_j)$. Wanneer er gegeven $\tilde{X}_i$ en $\tilde{Y}_j$ iets te zeggen valt over de vergelijking van X_i en Y_j, dan levert deze vergelijking een bijdrage aan de toetsingsgrootheid:
Voor het paar (X_i,Y_j) hebben we de volgende mogelijkheden:

(i) $\tilde{X}_i < \tilde{Y}_j$ en beide ongecensureerd $(\Delta_i=1,E_j=1) \Rightarrow X_i < Y_j \Rightarrow U_{ij} \overset{\text{def}}{=} -1$

(ii) $\tilde{X}_i \leq \tilde{Y}_j$ en $\tilde{X}_i$ ongecensureerd en $\tilde{Y}_j$ gecensureerd $(\Delta_i=1,E_j=0) \Rightarrow$
$X_i < Y_j \Rightarrow U_{ij} \overset{\text{def}}{=} -1$

(iii) $\tilde{X}_i > \tilde{Y}_j$ en beide ongecensureerd $(\Delta_i=1,E_j=1) \Rightarrow X_i > Y_j \Rightarrow U_{ij} \overset{\text{def}}{=} 1$

(iv) $\tilde{X}_i \geq \tilde{Y}_j$ en $\tilde{X}_i$ gecensureerd en $\tilde{Y}_j$ ongecensureerd ($\Delta_i=0, E_j=1$) $\Rightarrow$
$X_i > Y_j \Rightarrow U_{ij} \overset{\text{def}}{=} 1$

(v) In de overige gevallen valt er niets te zeggen over de grootte van X_i en Y_j ten opzichte van elkaar of $X_i = Y_j \Rightarrow U_{ij} \overset{\text{def}}{=} 0$.

<u>DEFINITIE 1.2.1</u>. De toetsingsgrootheid van de twee-steekproeven toets van Gehan wordt gedefinieerd als:

$$W = \sum_{i=1}^{m} \sum_{j=1}^{n} U_{ij} .$$

De berekeningen die Gehan geeft in zijn artikel, zijn nogal bewerkelijk. Ze kunnen echter aanzienlijk vereenvoudigd worden wanneer we de toetsingsgrootheid op een andere manier construeren. MANTEL [18] doet dit als volgt:

Voeg beide steekproeven samen tot een gecombineerde steekproef $\tilde{Z}_1, \tilde{Z}_2, \ldots, \tilde{Z}_{m+n}$. Voor zover we dit op grond van wat bekend is kunnen doen, vergelijken we de waarden in de gecombineerde steekproef. Definieer:

$$W_i \overset{\text{def}}{=} \#\{j : Z_i > Z_j,\ j \neq i\} - \#\{j : Z_i < Z_j,\ j \neq i\}$$

$$b_i \overset{\text{def}}{=} \begin{cases} 1 \text{ als } \tilde{Z}_i \text{ in eerste steekproef} \\ 0 \text{ als } \tilde{Z}_i \text{ in tweede steekproef} \end{cases}$$

$i = 1,2,\ldots,m+n.$

In de volgende gevallen krijgen we dus een bijdrage aan W_i:

	$\tilde{Z}_i$ ongecensureerd		$\tilde{Z}_i$ gecensureerd	
	$\tilde{Z}_j$ ongecensureerd	$\tilde{Z}_j$ gecensureerd	$\tilde{Z}_j$ ongecensureerd	$\tilde{Z}_j$ gecensureerd.
$\tilde{Z}_i < \tilde{Z}_j$	-1	-1	0	0
$\tilde{Z}_i = \tilde{Z}_j$	0	-1	+1	0
$\tilde{Z}_i > \tilde{Z}_j$	+1	0	+1	0

Nu geldt:

$$W = \sum_{i=1}^{m+n} b_i W_i .$$

Mantel geeft ook een schema om op een snelle wijze de waarden van de W_i's te bepalen:

STAP 1: Orden de gecombineerde steekproef in oplopende grootte (Daarbij geldt $\tilde{Z}_i \prec \tilde{Z}_j$ als $\tilde{Z}_i = \tilde{Z}_j$, $\tilde{Z}_i$ ongecensureerd, $\tilde{Z}_j$ gecensureerd).

STAP 2: Nummer de waarnemingen, daarbij gecens. waarnemingen overslaand.

STAP 3: Geef de gecens. waarnemingen een nummer dat 1 hoger is dan de ervoor liggende ongecens. waarneming.

STAP 4: Geef alle ongecens. waarnemingen in een knoop het laagst voorkomende rangnummer.

Dit geeft voor elke waarneming een getal R_1.

STAP 5: Nummer alle waarnemingen van rechts naar links.

STAP 6: Geef alle ongecens. waarnemingen in een knoop het laagst voorkomende rangnummer.

STAP 7: Geef alle gecens. waarnemingen rangnummer 1.

Dit geeft voor elke waarneming een getal R_2.

Voor iedere waarneming is de waarde van W_i nu gelijk aan $R_1 - R_2$.

We zullen de bovenstaande berekeningen illustreren aan de hand van een voorbeeld:

<u>VOORBEELD 1.2.2.</u>

We hebben twee steekproeven waarvan we de gelijkheid van de onderliggende verdeling willen toetsen. We nemen aan dat in beide gevallen de censurering dezelfde is.

S_1: 7, 8*, 8*, 11*, 3*, 12*, 3, 13*, 11*, 4, 2*, 4*, 2*, 12, 9*

S_2: 1, 6*, 4, 1, 1*, 1*, 6, 7, 3*, 3, 2, 5, 1, 3, 1*

Op de volgende bladzijde staat de berekening van W:

Onder de nulhypothese en voorwaardelijk op het gerealiseerde patroon P van W_i's kunnen we een aantal dingen zeggen over de toetsingsgrootheid W. Onder deze voorwaarden kunnen we immers de kansverdeling van W bepalen door alle permutaties van $b_1,\dots,b_{m+n}$ te nemen. Dit kunnen we alleen op correcte wijze toepassen indien we geval (iii) (identieke censurering) hebben en bovendien de censurering in beide steekproeven gelijk is.

Het volgende model kan dan gebruikt worden:

Zij $(B_1,\dots,B_{m+n})$ een stochastische vector die met kans $\frac{1}{\binom{m+n}{n}}$ de $\binom{m+n}{n}$

b_i	0	0	0	0	0	0	0	1	1	1	0	0	1	0	1	0	1	0	0	0	1	0	1	1	1	1	1	1	1	1
STAP 1:	1	1	1	1*	1*	1*	2	2*	2*	3	3	3	3*	3*	4	4	4*	5	6	6*	7	7	8*	8*	9*	11*	11*	12	12*	13*
STAP 2:	1	2	3				4			5	6	7			8	9		10	11		12	13						14		
STAP 3:				4	4	4		5	5				8	8			10			12			14	14	14	14	14		15	15
STAP 4:	1	1	1							5	5	5			8	8					12	12								
R_1:	1	1	1	4	4	4	4	5	5	5	5	5	8	8	8	8	10	10	11	12	12	12	14	14	14	14	14	14	15	15
STAP 5:	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
STAP 6:	28	28	28							19	19	19			15	15					9	9								
STAP 7:				1	1	1		1	1				1	1			1			1			1	1	1	1	1		1	1
R_2:	28	28	28	1	1	1	24	1	1	19	19	19	1	1	15	15	1	13	12	1	9	9	1	1	1	1	1	3	1	1
w_i:	-27	-27	-27	3	3	3	-20	4	4	-14	-14	-14	7	7	-7	-7	9	-3	-1	11	3	3	13	13	13	13	13	11	14	14

$$w = \sum_{i=1}^{m+n} b_i w_i = 110$$

onderling onderscheidbare permutaties van de vector $(b_1,\ldots,b_{m+n})$ aanneemt, waarbij $b_1=b_2=\ldots=b_m=1$, $b_{m+1}=b_{m+2}=\ldots=b_{m+n}=0$. Onder de voorwaarden geldt dan:

$$\{W|H_0,\mathcal{P}\} = \sum_{i=1}^{m+n} B_i w_i.$$

<u>LEMMA 1.2.3.</u>

$$E\{W|H_0,\mathcal{P}\} = 0.$$

<u>BEWIJS.</u>

$$E\{W|H_0,\mathcal{P}\} = \sum_{i=1}^{m+n} w_i \; EB_i = \frac{m}{m+n}\sum_{i=1}^{m+n} w_i = 0. \qquad \Box$$

<u>LEMMA 1.2.4.</u>

$$\mathrm{var}\{W|H_0,\mathcal{P}\} = \frac{mn}{(m+n)(m+n-1)}\sum_{i=1}^{m+n} w_i^2.$$

<u>BEWIJS.</u>

$$\mathrm{var}\{W|H_0,\mathcal{P}\} = E\{W^2|H_0,\mathcal{P}\} = E\{\sum_{i=1}^{m+n} B_i w_i\}^2 =$$

$$= \sum_{i=1}^{m+n} E\; B_i^2 w_i^2 + \sum_{i\neq j}^{m+n} E\; B_i B_j w_i w_j.$$

Aangezien

$$EB_i^2 = \frac{m}{m+n} \text{ en } EB_iB_j = \frac{m(m-1)}{(m+n)(m+n-1)}$$

volgt:

$$\mathrm{var}\{W|H_0,\mathcal{P}\} = \frac{m}{m+n}\sum_{i=1}^{m+n} w_i^2 + \frac{m(m-1)}{(m+n)(m+n-1)}\sum_{i\neq j}^{m+n} w_i w_j =$$

$$= \frac{m}{(m+n)}\left\{\left(1-\frac{(m-1)}{(m+n-1)}\right)\sum_{i=1}^{m+n} w_i^2 + \frac{(m-1)}{(m+n-1)}\left(\sum_{i=1}^{m+n} w_i\right)^2\right\} =$$

$$= \frac{mn}{(m+n)(m+n-1)}\sum_{i=1}^{m+n} w_i^2. \qquad \Box$$

OPMERKING 1.2.5. In het geval er geen censurering en knopen optreden, geldt voor de geordende steekproef: $w_i = (i-1)-(m+n-i) = 2i-(m+n+1)$, zodat we dan voor de variantie vinden:

$$\mathrm{var}\{W|H_0,P\} = \frac{mn}{(m+n)(m+n-1)} \sum_{i=1}^{m+n} (2i-m-n-1)^2 = \frac{1}{3}\, mn(m+n+1).$$

Dit is de variantie van de toetsingsgrootheid W van de bekende twee-steekproeven toets van Wilcoxon; W is daarbij gedefinieerd als

$$W \overset{\text{def}}{=} \sum_{i=1}^{m} \sum_{j=1}^{n} U_{ij} \qquad \text{met} \qquad U_{ij} \overset{\text{def}}{=} \begin{cases} 1 & \text{als } X_i > Y_j \\ 0 & \text{als } X_i = Y_j \\ -1 & \text{als } X_i < Y_j \end{cases}.$$

Voor grote waarden is de verdeling van de toetsingsgrootheid van Gehan niet eenvoudig te bepalen. Daarom is het prettig voor die gevallen een normale benadering te kunnen toepassen.

LEMMA 1.2.6. *Indien is voldaan aan:*

(a) $$\lim_{\substack{m\to\infty \\ n\to\infty}} \frac{m}{m+n} = \rho, \qquad \text{met } 0 < \rho < 1$$

(b) $w_1, w_2, \ldots, w_{m+n}$ *zijn zodanig, dat geldt:*

$$\frac{\sum_{i=1}^{m+n} w_i^r/(m+n)}{\left(\sum_{i=1}^{m+n} w_i^2/(m+n)\right)^{r/2}} = o((m+n)^{r/2-1}), \qquad (m+n) \to \infty, \qquad \text{voor } r = 3,4,\ldots$$

dan convergeert

$$(W|H_0,P)/\mathrm{var}\{W|H_0,P\}^{\frac{1}{2}}$$

voor $m,n \to \infty$ *naar een standaard normale verdeling.*

BEWIJS. In Lemma 1.2.7. blijkt dat onder de daar gestelde voorwaarden aan voorwaarde (b) is voldaan. We maken gebruik van een stelling van WALD-WOLFOWITZ-NOETHER-HOEFFDING (zie bijv. PURI & SEN [24], p.73, waar ook een sterkere voorwaarde staat die makkelijker is te verifiëren in de praktijk): Indien een toetsingsgrootheid geschreven kan worden in de vorm van een "lineaire permutatietoetsingsgrootheid"

$$L = \sum_{i=1}^{N} c_i X_i ,$$

waarin $X_1,\ldots,X_N$ als waarden de mogelijke permutaties van N reële getallen $a_1,\ldots,a_N$ aannemen, en bovendien is voldaan aan

$$\text{(I)}\qquad \frac{\sum_{i=1}^{N} (c_i-\bar{c})^r/N}{\left(\sum_{i=1}^{N} (c_i-\bar{c})^2/N\right)^{r/2}} = O(1),\ N \to \infty;\ \text{waarin } \bar{c} = \frac{1}{N}\sum_{i=1}^{N} c_i ;$$

$$\text{(II)}\qquad \frac{\sum_{i=1}^{N} (a_i-\bar{a})^r/N}{\left(\sum_{i=1}^{N} (a_i-\bar{a})^2/N\right)^{r/2}} = o(N^{r/2-1}),\ N \to \infty,\quad \text{waarin } \bar{a} = \frac{1}{N}\sum_{i=1}^{N} a_i ,$$

voor $r = 3,4,\ldots$.
Dan is $(L-E(L))/\mathrm{var}(L)^{1/2}$ asymptotisch standaard normaal verdeeld.

We nemen nu: $N = m+n$, $(c_1,\ldots,c_N) = (b_1,\ldots,b_{m+n})$ en $(a_1,\ldots,a_N) = (w_1,\ldots,w_{m+n})$. Inderdaad kan W nu geschreven worden in de vorm $W = \sum_{i=1}^{m+n} c_i X_i$.
Rest nog het controleren van voorwaarden I en II:

(I): Met behulp van (a) volgt:

$$\lim_{m,n\to\infty} \frac{\sum_{i=1}^{m+n} (b_i-\bar{b})^2/(m+n)}{\left(\sum_{i=1}^{m+n} (b_i-\bar{b})^2/(m+n)\right)^{r/2}} = \lim_{m,n\to\infty} \frac{\{m(\frac{n}{m+n})^r+n(\frac{-m}{m+n})^2\}/(m+n)}{\{m(\frac{n}{m+n})^2+n(\frac{-m}{m+n})^2\}^{r/2}/(m+n)^{r/2}} =$$

$$= \lim_{m,n\to\infty} \frac{(\frac{m}{m+n})(\frac{n}{m+n})^r+(\frac{n}{m+n})(\frac{m}{m+n})^r(-1)^r}{\{(\frac{m}{m+n})(\frac{n}{m+n})^2+(\frac{n}{m+n})(\frac{m}{m+n})^2\}^{r/2}} =$$

$$= \frac{\rho(1-\rho)^r+(1-\rho)\rho^r(-1)^r}{\{\rho(1-\rho)^2+(1-\rho)\rho^2\}^{r/2}} = \frac{(1-\rho)^{r-1}-(-\rho)^{r-1}}{\rho(1-\rho)^{r/2-1}} .$$

(II): Dit volgt direct uit (b), aangezien $\bar{w} = \frac{1}{m+n}\sum_{i=1}^{m+n} w_i = 0$. □

Tot besluit van de behandeling van deze toets bewijzen we nog dat hij asymptotisch onderscheidend is:

LEMMA 1.2.7. *Bij gelijke censurering en continue verdelingsfuncties* F *en* G *in beide steekproeven is de voorwaardelijke tweezijdige twee-steekproeven toets van Gehan asymptotisch onderscheidend tegen alternatieven van de vorm* $F(x) < G(x)\ \forall x$ *of* $F(x) > G(x)\ \forall x$, *indien* $m,n \to \infty$ *zodanig dat* $\lim_{m,n\to\infty} \frac{m}{m+n} = \rho$ *met* $0 < \rho < 1$.

BEWIJS. Definieer oneindige rijen $X_1, X_2, X_3, \ldots$, $U_1, U_2, U_3, \ldots, Y_1, Y_2, Y_3, \ldots$ en $V_1, V_2, V_3, \ldots$ op dezelfde kansruimte en bekijk de toetsingsgrootheid van Gehan gebaseerd op de eerste m waarnemingen van de eerste en tweede rij en de eerste n waarnemingen van de derde en vierde rij. Voer de volgende notatie in:

$$F_m^0(z) = \#\{i : \tilde{X}_i \le z,\ \Delta_i = 1\}/m$$

$$\tilde{F}_m(z) = \#\{i : \tilde{X}_i \le z\}/m$$

Analoog G_n^0 en $\tilde{G}_n$ voor de tweede steekproef.

$$H_{m,n}^0(z) = (mF_m^0(z)+nG_n^0(z))/(m+n)$$

$$\tilde{H}_{m,n}(z) = (m\tilde{F}_m(z)+n\tilde{G}_n(z))/(m+n).$$

Wegens de stelling van Glivenko-Cantelli geldt voor $m,n \to \infty$:

$$\left.\begin{array}{l} F_m^0(z) \to F^0(z) \overset{\text{def}}{=} \int_0^z (1-J(x))\,dF(x) \\ \tilde{F}_m(z) \to \tilde{F}(z) \overset{\text{def}}{=} 1-(1-F(z))(1-J(z)) \\ G_n^0(z) \to G^0(z) \overset{\text{def}}{=} \int_0^z (1-K(x))\,dG(x) \\ \tilde{G}_n(z) \to \tilde{G}(z) \overset{\text{def}}{=} 1-(1-G(z))(1-K(z)) \\ H_{m,n}^0(z) \to \rho F^0(z) + (1-\rho)G^0(z) \\ \tilde{H}_{m,n}(z) \to \rho\tilde{F}(z) + (1-\rho)\tilde{G}(z) \end{array}\right\} \text{ uniform in } z \text{ met kans } 1.$$

We kunnen de toetsingsgrootheid in de volgende vorm schrijven:

$$W = mn\{\int_0^\infty G_n^0(z)\,d\tilde{F}_m(z) - \int_0^\infty F_m^0(z)\,d\tilde{G}_n(z)\}.$$

Onder de voorwaarden van het lemma geldt $J = K$ zodat

$$\frac{W}{mn} \to \int_0^\infty G^0(z)d\tilde{F}(z) - \int_0^\infty F^0(z)d\tilde{G}(z), \quad \text{met kans 1.}$$

Nu volgt:

$$\int_0^\infty G^0(z)d\tilde{F}(z) - \int_0^\infty F^0(z)d\tilde{G}(z) =$$

$$= [G^0(z)\tilde{F}(z)]_0^\infty - \int_0^\infty \tilde{F}(z)dG^0(z) - [F^0(z)\tilde{G}(z)]_0^\infty + \int_0^\infty \tilde{G}(z)dF^0(z) =$$

$$= \int_0^\infty dG^0(z) - \int_0^\infty \tilde{F}(z)dG^0(z) - \int_0^\infty dF^0(z) + \int_0^\infty \tilde{G}(z)dF^0(z) =$$

$$= \int_0^\infty (1-\tilde{F}(z))dG^0(z) - \int_0^\infty (1-\tilde{G}(z))dF^0(z) =$$

$$= \int_0^\infty (1-J(z))^2(1-F(z))dG(z) - \int_0^\infty (1-J(z))^2(1-G(z))dF(z) =$$

$$= \int_0^\infty (1-J(z))^2(G(z)-F(z))dF(z) + \int_0^\infty (1-J(z))^2(1-F(z))d(G(z)-F(z)) =$$

$$= \int_0^\infty (1-J(z))^2(G(z)-F(z))dF(z) + \int_0^\infty (G(z)-F(z))d(1-(1-F(z))(1-J(z))^2),$$

zodat W onder het alternatief $F < G$ of $F > G$ met kans 1 naar $+\infty$ resp. $-\infty$ convergeert.

De W_i's kunnen we schrijven in de volgende vorm:

$$\frac{W_i}{m+n} = H^0_{m,n}(\tilde{Z}_i) - \frac{1}{m+n} - (1-\tilde{H}_{m,n}(\tilde{Z}_i))\Gamma_i,$$

met $\Gamma_i = \Delta_i$ resp. E_i.
Dus:

$$\sum_{i=1}^{m+n} W_i^2/(m+n)^2 = (m+n)\{\int_0^\infty (H^0_{m,n}(z) - \frac{1}{m+n})^2 d\tilde{H}_{m,n}(z) + \int_0^\infty (1-\tilde{H}_{m,n}(z))^2 dH^0_{m,n}(z) +$$

$$- 2\int_0^\infty (H^0_{m,n}(z) - \frac{1}{m+n})(1-\tilde{H}_{m,n}(z))dH^0_{m,n}(z)\}$$

zodat

$$\lim_{m+n\to\infty} \frac{\mathrm{var}\{W|H_0,P\}}{(m+n)^3} = \rho(1-\rho)\int_0^\infty (H^0(z)-1+\tilde{H}(z))^2 dH^0(z)+\int_0^\infty (H^0(z))^2 d(\tilde{H}(z)-H^0(z)) > 0,$$

met kans 1.

m.a.w.
$$\frac{W}{\mathrm{var}\{W|H_0,P\}^{1/2}} \to \pm\infty, \quad \text{als } m+n\to\infty \text{ met kans 1.}$$

Op dezelfde wijze kunnen we aantonen dat met kans 1 aan voorwaarde (b) van lemma 1.2.6. is voldaan met het gevolg dat het acceptatiegebied voor de nulhypothese begrensd is als $m + n \to \infty$.

Concluderend kunnen we zeggen dat onder de alternatieve hypothese de toetsingsgrootheid met kans 1 niet in het acceptatiegebied komt te liggen.

Teruggaand naar voorbeeld 1.2.2., zou de gestandaardiseerde toetsingsgrootheid een waarde gelijk aan 3,05 krijgen, die bij een onbetrouwbaarheid van 5% buiten het acceptatiegebied (-1,96,+1,96) zou komen te liggen. Aannemende dat de normale benadering toegepast kan worden, zouden we de nulhypothese moeten verwerpen.

1.3. DE PRODUCT-LIMIET SCHATTER

Alvorens verder te gaan met de twee overige methoden, willen we eerst een techniek bespreken om de overlevingsfunctie te schatten. We maken daarbij gebruik van de bekende notatie. Er worden op dit moment nog geen eisen gesteld aan de verdelingsfuncties F van de wachttijden $X_1,\ldots,X_m$ en J van de censureringsvariabelen $U_1,\ldots,U_m$. Wel maken we de volgende veronderstellingen:

(i) $X_1,\ldots,X_m$ zijn o.o. identiek verdeelde stochastische variabelen.
(ii) $U_1,\ldots,U_m$ zijn o.o. identiek verdeelde stochastische variabelen.
(iii) De X_i's en U_i's zijn o.o. .

Nu geldt voor de verdelingsfunctie $\tilde{F}$ van de $\tilde{X}_i$'s

$$\tilde{F}(x) = P(\tilde{X}_i\le x) = 1-P(\tilde{X}_i>x) =$$

$$= 1-P(X_i>x)P(U_i>x) = 1-(1-F(x))(1-J(x)).$$

Verder geldt op grond van de stelling van Fubini:

$$F^0(x) = P(\tilde{X}_i \le x, \Delta_i = 1) = P(X_i \le x, \Delta_i = 1)$$

$$= \int_0^x P(U_i \ge t)\,dF(t) = \int_0^x (1 - J(t-))\,dF(t).$$

We willen S nu schatten in het punt x. Daartoe verdelen we het interval $(0,x]$ in k deelintervallen $(\tau_{i-1},\tau_i]$, met $0 = \tau_0 < \tau_1 < \ldots < \tau_k = x$. Zij q_i nu de kans dat de wachttijd van de i-de gebeurtenis valt in het interval $(\tau_{i-1},\tau_i]$ onder de voorwaarde dat de wachttijd groter is dan τ_{i-1}:

$$q_i \overset{\text{def}}{=} P(X_i \in (\tau_{i-1},\tau_i] \mid X_i > \tau_{i-1}) = \frac{F(\tau_i) - F(\tau_{i-1})}{1 - F(\tau_i)} =$$

$$= \frac{S(\tau_{i-1}) - S(\tau_i)}{S(\tau_i)}$$

$$p_i \overset{\text{def}}{=} 1 - q_i.$$

Dan geldt m.b.v. de productregel:

$$P(X>x) = 1 - F(\tau_k) =$$

$$= P(X \notin (\tau_0,\tau_1] \mid X > \tau_0)\, P(X \notin (\tau_1,\tau_2] \mid X > \tau_1) \ldots P(X \notin (\tau_{k-1},\tau_k] \mid X > \tau_{k-1}) =$$

$$= \prod_{i=1}^{k} (1 - q_i) = \prod_{i=1}^{k} p_i.$$

Het idee is nu om op basis van een intervalverdeling de p_i's te schatten en daaruit een schatter voor S te construeren, een schatter die beter wordt naarmate we de indeling van het interval verfijnen (zie KAPLAN & MEIER [16]). Wanneer de indeling van het interval voldoende fijn is en de onderliggende verdeling continu, dan bevat ieder deelinterval hoogstens één waarneming. Laten $\tilde{X}_{(1)},\ldots,\tilde{X}_{(m)}$ de geordende waarnemingen zijn met $\Delta_{(1)},\ldots,\Delta_{(m)}$ de bijbehorende censuur-indicatoren. Dan schatten we voor het deelinterval $(\tau_{j-1},\tau_j]$ q_j met

$$\hat{Q}_j \overset{\text{def}}{=} \begin{cases} \dfrac{1}{(m-i+1)}, & \text{als } \tilde{X}_{(i)} \in (\tau_{j-1},\tau_j] \text{ en } \Delta_{(i)} = 1 \\ 0, & \text{overigens} \end{cases}$$

<u>DEFINITIE 1.3.1</u>. De product-limiet schatter S_m voor de overlevingsfunctie S wordt gedefinieerd door:

$$S_m(x) = \begin{cases} \prod_{\{i:\tilde{X}_{(i)} \le x, \Delta_{(i)}=1\}} \left(1-\frac{1}{m-i+1}\right) = \prod_{\{i:\tilde{X}_{(i)} \le x\}} \left(\frac{m-i}{m-i+1}\right)^{\Delta(i)} & \text{als } x < \tilde{X}_{(m)} \\ 0, \quad \text{als } x \ge \tilde{X}_{(m)} & \end{cases}$$

Deze definitie kan ook gebruikt worden in het geval er knopen optreden. Wanneer de grootste waarneming $\tilde{X}_{(m)}$ ongecensureerd is dan geldt $S_m(x) = 0$ voor $x \ge \tilde{X}_{(m)}$ in de eerste uitdrukking. Als $\tilde{X}_{(m_0)}$ de grootste ongecensureerde waarneming is met $m_0 < m$ dan is $S_m(x) > 0$ voor $x \ge \tilde{X}_{(m_0)}$ in de eerste uitdrukking. We kunnen dus niet volstaan met de eerste uitdrukking, aangezien dit in het algemeen geen overlevingsfunctie is.

De PL-schatters voor de overlevingsfunctie zijn consistent en hebben een kleine onzuiverheid. We bewijzen nu dat de PL-schatter een meest aannemelijke schatter is:

LEMMA 1.3.2. *De PL-schatter is (binnen de klasse van discrete verdelingen) voorwaardelijk op de waarden van de censurerende stochasten* $U_1,\ldots,U_m$ *een meest aannemelijke schatter voor de overlevingsfunctie.*

BEWIJS. Laten $X_1^0,\ldots,X_k^0$ de geordende ongecensureerde waarnemingen zijn en zij $X_0^0 = 0$ en $X_{k+1}^0 = \infty$. Laten $U_i^{(j)}$, de gecensureerde waarnemingen in het interval $[X_j^0, X_{j+1}^0)$ zijn, $i = 1,\ldots,N_j$, $j = 0,\ldots,k$. Zij S een willekeurige overlevingsfunctie. Dan is de aannemelijkheidsfunctie, gegeven $U_i^{(j)} = u_i^{(j)}$ gelijk aan:

$$\prod_{i=1}^{N_0} S(u_i^{(0)}) \prod_{j=1}^{k} \{(S(X_j^0-) - S(X_j^0)) \prod_{i=1}^{N_j} S(u_i^{(j)})\}.$$

We moeten nu die S vinden waarvoor deze functie gemaximaliseerd wordt. Dit houdt in dat $S(u_i^{(j)})$ en $S(X_j^0-)$ zo groot mogelijk en $S(X_j^0)$ zo klein mogelijk moeten worden, onder de voorwaarde dat S een overlevingsfunctie blijft. Kies S zodanig dat

$$S(u_i^{(0)}) = 1 \quad \text{en} \quad S(X_j^0) = S(u_i^{(j)}) = S(X_{j+1}^0-).$$

Definieer nu:

$$R_j \overset{\text{def}}{=} S(X_j^0),\ R_0 = 0,\ P_j \overset{\text{def}}{=} 1-Q_j = R_j/R_{j-1}, \qquad j = 1,\ldots,k.$$

Dan gaat de aannemelijkheidsfunctie door substitutie over in,

$$\prod_{j=1}^{k} (R_{j-1}-R_j)R_j^{N_j}.$$

Aangezien

$$R_j = \prod_{i=1}^{j} P_i \quad \text{en} \quad R_{j-1} - R_j = \prod_{i=1}^{j-1} P_i Q_j,$$

krijgen we voor de aannemelijkheidsfunctie:

$$\prod_{j=1}^{k} P_j^{M_j-1} Q_j, \qquad \text{met} \qquad M_j = m - \sum_{i=0}^{j-1} N_j - (j-1).$$

In deze vorm zal elke factor met vaste index j worden gemaximaliseerd door de schatting

$$\hat{P}_j = \frac{M_j-1}{M_j}, \quad \hat{Q}_j = 1 - \hat{P}_j, \quad \hat{R}_j = \prod_{i=1}^{j} \hat{P}_i,$$

zodat de meest aannemelijke schatter gelijk is aan:

$$S_m(x) = \prod_{\{i:X_i^0 \le x\}} S_m(X_i^0)/S_m(X_{i-1}^0) = \prod_{\{i:X_i^0 \le x\}} \hat{R}_i/\hat{R}_{i-1} = \prod_{\{i:X_i^0 \le x\}} \hat{P}_i$$

hetgeen hetzelfde is als de PL-schatter. □

1.4. DE TOETS VAN EFRON

EFRON [13] construeerde een zodanige toets dat, naar analogie van de toets van Wilcoxon, bij verwerping van de nulhypothese, de waarde van de toetsingsgrootheid een maat voor de afwijking van de nulhypothese is, dus bijvoorbeeld een schatting voor $P(X>Y)$. De toetsingsgrootheid zoals geconstrueerd door Gehan heeft deze eigenschap niet, aangezien in het geval $F \neq G$ de verwachting van de toetsingsgrootheid nog afhangt van de verdelingen van censureringsvariabelen U en V. In deze paragraaf zullen we de constructie van de toets en enige eigenschappen bekijken.

Laten S_m en T_n de PL-schatters voor de eerste en tweede steekproef zijn. De eerste steekproef wordt weer gevormd door $\tilde{X}_1,\ldots,\tilde{X}_m$ met $\tilde{X}_i = \min\{X_i,U_i\}$, $i = 1,2,\ldots,m$ en de tweede door $\tilde{Y}_1,\ldots,\tilde{Y}_n$ met $\tilde{Y}_i = \min\{Y_i,V_i\}$, $i = 1,2,\ldots,n$.

DEFINITIE 1.4.1. De toetsingsgrootheid van de twee-steekproeven toets van Efron is gelijk aan

$$W \stackrel{\text{def}}{=} - \int_0^\infty S_m(x)\,dT_n(x).$$

Aangezien met behulp van Fubini geldt:

$$- \int_0^\infty S(x)\,dT(x) = P(X>Y),$$

kunnen we stellen dat W een schatter is voor $P(X>Y)$.

LEMMA 1.4.2. *Indien* $m,n \to \infty$ *zodanig, dat*

$$\lim_{m,n\to\infty} \frac{m}{m+n} = \rho \quad \textit{met} \quad 0 < \rho < 1,$$

dan geldt bij Model-I-censurering en continue F, G, J *en* K *onder* H_0*; dat*

$$\frac{\sqrt{m+n}\,(W-P(X>Y))}{\sqrt{\sigma_1^2/\rho+\sigma_2^2/(1-\rho)}}$$

asymptotisch standaard normaal verdeeld is, als

$$\sigma_1^2 = \frac{1}{4}\int_0^1 \frac{t^2\,dt}{1-J(S^{-1}(t))} = \frac{1}{4}\int_0^1 \frac{t^3\,dt}{\tilde{S}(S^{-1}(t))} < \infty$$

en

$$\sigma_2^2 = \frac{1}{4}\int_0^1 \frac{t^2\,dt}{1-K(T^{-1}(t))} = \frac{1}{4}\int_0^1 \frac{t^3\,dt}{\tilde{T}(T^{-1}(t))} < \infty$$

waarbij

$$\tilde{S}(x) = P(\tilde{X}>x) \quad \textit{en} \quad \tilde{T}(x) = P(\tilde{Y}>x).$$

BEWIJS. Dit is het resultaat van Efron, die volstaat met het geven van onderstaand schetsmatig bewijs. Het is nog onduidelijk of zonder verdere condities dit bewijs inderdaad juist is. Onder sterkere voorwaarden zal in hoofdstuk 3 een ander bewijs gegeven worden.

$$- W = \int_0^\infty S_m(x)dT_n(x) =$$

$$= \int_0^\infty S(x)dT(x) + \int_0^\infty (S_m(x)-S(x))dT(x) + \int_0^\infty S(x)d(T_n(x)-T(x)) +$$

$$\int_0^\infty (S_m(x)-S(x))d(T_n(x)-T(x)).$$

Partieel integreren van de derde term levert:

$$-W = \int_0^\infty S(x)dT(x) + \int_0^\infty (S_m(x)-S(x))dT(x) - \int_0^\infty (T_n(x)-T(x))dS(x) +$$

$$\int_0^\infty (S_m(x)-S(x))d(T_n(x)-T(x)).$$

We kunnen $\sqrt{m}(S_m(x)-S(x))$ opvatten als een stochastisch proces in x. Als we dat doen dan blijkt dit proces voor $m \to \infty$ te convergeren naar een normaal proces met verwachting 0 en covariantie

$$\Gamma(x,y) = S(x)S(y) \int_0^x \frac{-dS(u)}{\tilde{S}(u)S(u)}, \qquad x \le y.$$

Merk op dat de eerste term in de ontwikkeling van W een constante is. Voor de vierde term geldt:

$$\int_0^\infty (S_m(x)-S(x))d(T_n(x)-T(x)) =$$

$$= \frac{1}{\sqrt{mn}} \int_0^\infty \sqrt{m}(S_m(x)-S(x))d(\sqrt{n}(T_n(x)-T(x))) = o_P((m+n)^{-\frac{1}{2}}), \quad m,n \to \infty.$$

Verder geldt dat

$$\sqrt{m} \int_0^\infty (S_m(x)-S(x))dT(x)$$

convergeert naar een normale verdeling met verwachting 0 en variantie

$$\int_0^\infty \int_0^\infty \Gamma(x,y)dT(x)dT(y) = -2 \int_0^\infty \int_x^\infty \int_0^x \frac{S(x)S(y)}{\tilde{S}(u)S(u)} dS(u)dT(y)dT(x).$$

Aangezien onder de nulhypothese geldt dat S = T, wordt deze laatste uitdrukking:

$$2\int_0^1\int_0^x\int_x^1 \frac{xy}{u\tilde{S}(S^{-1}(u))}\,du\,dy\,dx = 2\int_0^1\int_0^u\int_0^x \frac{xy}{u\tilde{S}(S^{-1}(u))}\,dy\,dx\,du =$$

$$= \frac{1}{4}\int_0^1 \frac{u^3du}{\tilde{S}(S^{-1}(u))} = \frac{1}{4}\int_0^1 \frac{u^2du}{1-J(S^{-1}(u))} = \sigma_1^2\ .$$

Voor de derde term vinden we op analoge wijze σ_2^2. Met behulp van de onafhankelijkheid van 2e en 3e term vindt men dan het gevraagde. □

Tot slot van deze paragraaf willen we de berekening van de toetsingsgrootheid illustreren aan de hand van het voorbeeld uit paragraaf 1.2.

Merk daarbij wel op dat de toets gebaseerd is op continue onderliggende verdelingen. Toch kunnen we de toetsingsgrootheid wel uitrekenen, daarbij gebruik makend van dezelfde formule voor de PL-schatter:

VOORBEELD 1.4.3. We hebben de volgende twee geordende steekproeven:

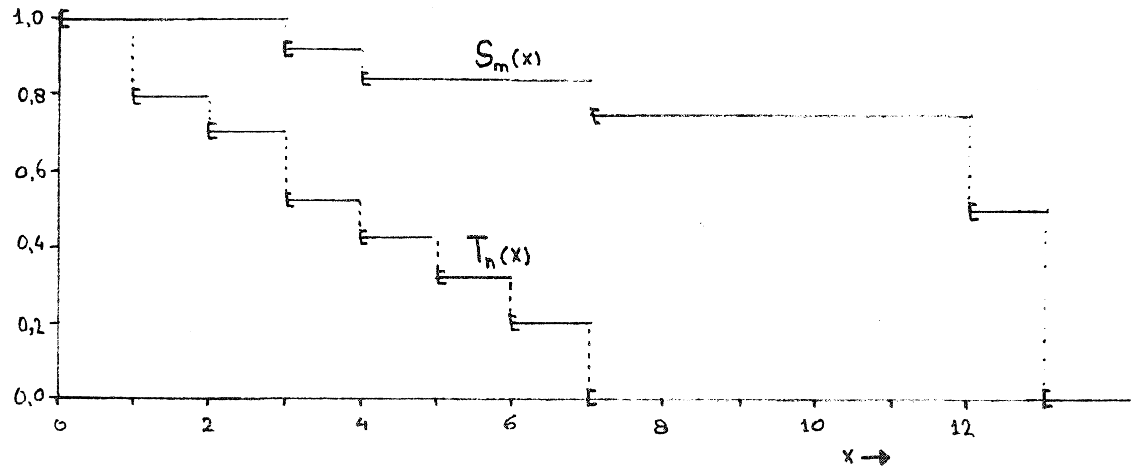
S_1: 2*, 2*, 3, 3*, 4, 4*, 7, 8*, 8*, 9*, 11*, 11*, 12, 12*, 13*

S_2: 1, 1, 1, 1*, 1*, 1*, 2, 3, 3, 3*, 4, 5, 6, 6*, 7

De product-limiet schatters S_m en T_n voor beide steekproeven staan in onderstaande tabel

SPRONGPUNT	S_m	T_n
	1,00	1,00
1	1,00	0,80
2	1,00	0,71
3	0,92	0,53
4	0,84	0,43
5	0,84	0,32
6	0,84	0,21
7	0,75	0,00
12	0,50	0,00
13	0,00	0,00

In een grafiek:



Uit de grafiek blijkt duidelijk dat $S_m(x) > T_n(x)$.
De waarde van de toetsingsgrootheid berekenen we als volgt:

$$W = \int_0^\infty S_m(x)dT_n(x) = \sum_{i=1}^{n} S_m(\tilde{Y}_i)[T_n(\tilde{Y}_i-)-T_n(\tilde{Y}_i)]E_i.$$

We vinden voor W 0,88 wat er op wijst dat objecten uit de eerste steekproef grotere wachttijden hebben dan objecten uit de tweede steekproef.

We kunnen ρ, σ_1^2 en σ_2^2 als volgt schatten:

$$\rho \text{ met } \frac{m}{m+n};$$

$$\sigma_1^2 \text{ met } \frac{1}{4}\int_0^1 \frac{t^3dt}{\tilde{S}_m(S_m^{-1}(t))} = \frac{1}{16}\sum_{i=1}^{m}[S_m(\tilde{X}_i-)^4 - S_m(\tilde{X}_i)^4]\Delta_i/\tilde{S}_m(\tilde{X}_i-);$$

$$\sigma_2^2 \text{ met } \frac{1}{4}\int_0^1 \frac{t^3dt}{\tilde{T}_m(T_n^{-1}(t))} = \frac{1}{16}\sum_{i=1}^{n}[T_n(\tilde{Y}_i-)^4-T_n(\tilde{Y}_i)^4]E_i/\tilde{T}_n(\tilde{Y}_i-)$$

Aannemende dat ook bij het optreden van knopen de gestandaardiseerde toetsingsgrootheid onder H_0 bij benadering standaard normaal verdeeld is, vinden we een waarde van 2,83 die buiten het acceptatiegebied (-1,96, +1,96) ligt bij een onbetrouwbaarheid van 5%. We moeten de nulhypothese dat de onderliggende wachttijdverdelingen gelijk zijn, dus verwerpen.

1.5. DE TOETS VAN COX (De log rank toets)

COX [7] bekijkt een meer algemene situatie waarin ook nog andere variabelen (covariabelen) optreden die de gang van zaken beïnvloeden. Men weet bijvoorbeeld dat bij een onderzoek naar het levensverlengende effect van geneesmiddelen voor kankerpatiënten factoren als leeftijd, oorspronkelijke gezondheidstoestand, enz. een sterk effect op het uiteindelijke resultaat kunnen hebben. Een manier om de invloed van die externe factoren te elimineren is het uitsplitsen van de te onderzoeken populatie in een aantal homogene deelpopulaties. Wanneer er echter veel covariabelen en weinig onderzoeksobjecten zijn, zullen maar weinig conclusies getrokken kunnen worden. Het blijkt dan noodzakelijk om de covariabelen als parameters in het model op te nemen. Men zou daarbij kunnen denken aan een lineair model, een model waarin de waarde van de te onderzoeken variabele een lineaire samenhang vertoont met de waarde van de covariabelen. De consequentie van een dergelijke modelveronderstelling is echter dat men er van uitgaat dat de covariabelen alleen effect hebben op de verwachting van de te onderzoeken variabele. In de meeste praktijksituaties is dit echter een onaanvaardbare veronderstelling. Een wel aanvaardbaar uitgangspunt in concrete levensduuronderzoekingen is dat het risico factoriseert in een factor, alleen afhankelijk van de tijd, en een factor, alleen afhankelijk van de covariabelen. We gaan het begrip risico in een mathematisch model onderbrengen en we definiëren daartoe het begrip uitvalsfunctie. ("risicofunctie" zou verwarrend kunnen werken). (Engels: "hazard rate"). Zij X een stochastische variabele die de wachttijd voorstelt. Men is vaak geïnteresseerd in de kans dat een wachttijd in een bepaald interval ligt, gegeven dat hij minstens zo groot is als de ondergrens van het interval:

$$P(x \le X < x+h \mid X \ge x).$$

<u>DEFINITIE 1.5.1</u>. De uitvalsfunctie $\lambda(x)$ wordt als volgt gedefinieerd:

(i) Voor (absoluut) continue verdelingen:

$$\lambda(x) = \lim_{h \downarrow 0} \frac{1}{h} P(x \le X < x+h) \mid X \ge x) = \frac{f(x)}{1-F(x)} = \frac{f(x)}{S(x)},$$

waarin f en F dichtheid resp. verdelingsfunctie voorstellen.

(ii) Voor discrete verdelingen:

$$\lambda(x) = \frac{P(X=x)}{1-F(x-)} = \frac{P(X=x)}{S(x-)}.$$

GEVOLG 1.5.2.

(i) Voor continue verdeling:

$$\frac{d \log S(x)}{dx} = \frac{S'(x)}{S(x)} = -\frac{f(x)}{S(x)} = -\lambda(x)$$

$$\Rightarrow \log S(x) = -\int_0^x \lambda(y)dy \Rightarrow S(x) = \exp\{-\int_0^x \lambda(y)dy\}$$

(ii) voor discrete verdeling (met geordende drager $\{x_1,x_2,\ldots\}$):

$$1 - \lambda(x_i) = 1 - \frac{P(X=x_i)}{S(x_i-)} = \frac{S(x_i)}{S(x_i-)} .$$

Kies x willekeurig en zij $k = \max\{i:x_i \le x\}$

$$\Rightarrow S(x) = S(x_k) = S(x_1-) \prod_{i=1}^{k} \frac{S(x_i)}{S(x_i-)} = \prod_{i=1}^{k} \frac{S(x_i)}{S(x_i-)} = \prod_{i=1}^{k} (1-\lambda(x_i)).$$

In een gemengde verdeling heeft men componenten van beide typen.

De uitvalsfunctie heeft een zeer natuurlijke plaats in de theorie der levensduurverdelingen. Voor ieder tijdstip waarop sterfte plaats vindt, wordt de kans daarop geschat op voor de hand liggende wijze en daarna wordt m.b.v. 1.5.2.(ii) de overlevingsfunctie geschat. De analogie met de PL-schatter is duidelijk.

We bouwen nu een regressie model, er daarbij van uitgaande dat X een continue verdeling heeft. Stel dat de waarde van X beïnvloed wordt door een vector van p covariabelen $w = (w_1,\ldots,w_p)'$. Het model van Cox houdt nu in dat de uitvalsfunctie (die een functie is van x en w) geschreven kan worden in de volgende vorm:

1.5.3. $\lambda(x;w) = h(w)\ \lambda_0(x)$,

waarin $h(w) = e^{w'\beta}$; $\beta = (\beta_1,\ldots,\beta_p)$ een vector van onbekende parameters en λ_0 een (onbekende) uitvalsfunctie bij de standaard waarde $w = 0$.

De in 1.5.3 beschreven vorm is gekozen voor het gemak. Het is natuurlijk ook mogelijk een analyse uit te voeren op grond van een andere modelspecificatie.

Uit 1.5.2.(i) en 1.5.3 volgt: (S is nu een functie van x en w):

$$S(x;w) = \exp\{-\int_0^x \lambda(y;w)dy\} = \exp\{-\int_0^x h(w)\lambda_0(y)dy\} =$$

$$= \exp\{-\int_0^x \lambda_0(y)dy\}^{h(w)} = S(x;0)^{h(w)}.$$

Dit definieert een z.g. Lehmann-familie van verdelingen.

Analyse van het model houdt ondermeer in: Schatten van β, toetsen van hypothesen omtrent β en schatten van S. COX [7] geeft allerlei intuïtieve argumenten om zijn methode aannemelijk te maken. De theorie van het twee-steekproeven geval zullen we later uitgebreider behandelen. De uitwerking voor het twee-steekproeven geval loopt verder als volgt:
Laat w 1-dimensionaal zijn met twee mogelijke waarden: 0 voor de eerste steekproef en 1 voor de tweede steekproef.
Voor de steekproeven geldt nu:

$$S(x) = S(x;0)$$

$$T(x) = S(x;1) = S(x;0)^{h(1)} = S(x;0)^{e^{\beta}}.$$

De nulhypothese van gelijke levensduur verdelingen houdt dus in: $\beta = 0$ en dat is de hypothese die we in dit geval gaan toetsen.

Later zal blijken dat de toets ook redelijk werkt voor de meer algemene alternatieven $F(x) < G(x)$ of $F(x) > G(x)$ $\forall x$, waarin $F(x)$ en $G(x)$ de verdelingsfuncties bij de eerste en tweede steekproef.

Zonder verlies van algemeenheid kunnen we voor de verdere berekeningen stellen dat $\tilde{Z}_{(1)}, \tilde{Z}_{(2)}, \ldots, \tilde{Z}_{(m+n)}$ de geordende gecombineerde steekproef is. Om de notatie wat te vereenvoudigen nemen we aan dat $Z_1^0, \ldots, Z_s^0$ de geordende waarden van de ongecensureerde waarnemingen zijn. Zij $R_i \overset{\text{def}}{=} \{j : \tilde{Z}_j \geq Z_i^0\}$. R_i wordt wel de i-de *uitvalsverzameling* genoemd. $(i = 1,2,\ldots,s)$.

Cox geeft een aantal heuristische argumenten waarom bij bepaalde typen censurering voorwaardelijk op R_i de kans dat het object met index k uit R_i ongecensureerd is met waarde Z_i^0, gelijk is aan:

$$\exp\{w_k\beta\} / \sum_{j\in R_i} \exp\{w_j\beta\}.$$

Iedere ongecensureerde waarneming draagt een dergelijke factor bij, zodat

L, de logaritme van de aannemelijkheidsfunctie, gelijk is aan:

$$L(\beta) = \sum_{i=1}^{s} W_i^0\beta - \sum_{i=1}^{s} \log\left[\sum_{j\in R_i} \exp\{w_j\beta\}\right],$$

waarin W_i^0 de waarden van w corresponderend met Z_i^0. Hieruit volgt:

$$U(\beta) \overset{\text{def}}{=} \frac{\partial L(\beta)}{\partial\beta} = \sum_{i=1}^{s} W_i^0 - \sum_{i=1}^{s} \left(\frac{\sum_{j\in R_i} w_j\exp\{w_j\beta\}}{\sum_{j\in R_i} \exp\{w_j\beta\}} \right);$$

$$D(\beta) \overset{\text{def}}{=} \frac{\partial^2 L(\beta)}{\partial\beta^2} =$$

$$= -\sum_{i=1}^{s} \left(\left(\frac{\sum_{j\in R_i} w_j^2\exp\{w_j\beta\}}{\sum_{j\in R_i} \exp\{w_j\beta\}} \right) - \left(\frac{\sum_{j\in R_i} w_j\exp\{w_j\beta\}}{\sum_{j\in R_i} \exp\{w_j\beta\}} \right)^2 \right)$$

Zij $\Pi_i \overset{\text{def}}{=}$ de fractie van objecten uit R_i waarvoor w = 1 en zij N_1 het aantal ongecensureerde waarnemingen waarvoor w = 1, dan

DEFINITIE 1.5.4.

$$U(0) = N_1 - \sum_{i=1}^{s} \Pi_i$$

$$-D(0) = \sum_{i=1}^{s} \Pi_i(1-\Pi_i).$$

Schatting van β blijkt niet zonder meer mogelijk te zijn. Dit kan alleen met numerieke technieken op een rekenautomaat.

We kunnen wel een asymptotisch normaal (0,1) verdeelde toetsingsgrootheid voor de hypothese H_0: β = 0 vinden; neem namelijk

$$\frac{U(0)}{\sqrt{-D(0)}}.$$

De asymptotische normaliteit zal later bewezen worden. Deze toetsingsgrootheid staat ook beschreven in THOMAS [25] en PETO & PETO [22] en wordt daar "log rank test" genoemd. In wezen bekijken we s kruistabellen, 1 voor iedere uitvalsverzameling, die er als volgt uitzien:

	$w = 0$	$w = 1$	
uitvallen op Z_i^0			1
niet uitvallen op Z_i^0			$\#R_i-1$
	$(1-\Pi_i)\#R_i$	$\Pi_i\#R_i$	$\#R_i$

Onder de nulhypothese en gegeven de randtotalen heeft het aantal ongecensureerde waarnemingen uit de tweede steekproef dat uitvalt op het tijdstip Z_i^0 verwachting Π_i en variantie $\Pi_i(1-\Pi_i)$. (Hokje rechtsboven)

We zullen het gebruik van deze toets illustreren aan de hand van voorbeeld 1.2.2. Daarbij zullen we de door Cox gegeven formules voor het geval van knopen gebruiken:

Zij $M_i \overset{\text{def}}{=} \#\{j: \tilde{Z}_j = Z_i^0,\ \Delta_j = 1$, resp. $E_j = 1\}$ en

$R_i \overset{\text{def}}{=} \#\{j: \tilde{Z}_j \geq Z_i^0\}$, dan geldt:

$$U(0) = N_1 - \sum_{i=1}^{s} M_i \Pi_i .$$

$$-D(0) = \sum_{i=1}^{s} \frac{M_i(R_i-M_i)}{R_i-1} \Pi_i(1-\Pi_i) .$$

<u>VOORBEELD 1.5.5.</u>

S_1: 2*, 2*, 3, 3*, 4, 4*, 7, 8*, 8*, 9*, 11*, 11*, 12, 12*, 13*. (w=0)

S_2: 1, 1, 1, 1*, 1*, 1*, 2, 3, 3, 3*, 4, 5, 6, 6*, 7. (w=1)

De berekeningen voor U(0) en -D(0) volgen hieronder: (op de volg. bladz.)

Als we aannemen dat de steekproef omvang groot genoeg is om de normale benadering toe te kunnen passen, dan is het acceptatiegebied bij een onbetrouwbaarheid van 5% gelijk aan (-1,96,+1,96). De gevonden waarde is gelijk aan

$$\frac{5{,}599}{\sqrt{2{,}531}} = 3{,}519$$

en dit ligt buiten het acceptatiegebied zodat we de nulhypothese moeten verwerpen.

X	M	w	M_i	R_i	Π_i	$M_i\Pi_i$	$\frac{M_i(R_i-M_i)}{R_i-1}\,\Pi_i(1-\Pi_i)$
1	3	1	3	30	0,500	1,500	0,698
1*	3	1					
2	1	1	1	24	0,375	0,375	0,234
2*	2	0					
3	1	0	3	21	0,381	1,143	0,637
	2	1					
3*	1	0					
	1	1					
4	1	0	2	16	0,313	0,625	0,401
	1	1					
4*	1	0					
5	1	1	1	13	0,308	0,308	0,213
6	1	1	1	12	0,250	0,250	0,188
6*	1	1					
7	1	0	2	10	0,100	0,200	0,160
	1	1					
8*	2	0					
9*	1	0					
11*	2	0					
12	1	0	1	3	0,000	0,000	0,000
12*	1	0					
13*	1	0					

	+ 4,401	2,531 +

$\Rightarrow$ $U(0) = N_1 - 4{,}401 = 10 - 4{,}401 = 5{,}599$ en $-D(0) = 2{,}531$.

We hebben nu drie verschillende toetsen bekeken en het is niet eenvoudig in te zien wat de relatieve merites van deze toetsen ten opzichte van elkaar zijn. Om nu een vergelijking mogelijk te maken gaan we in het volgende hoofdstuk eerst een stuk theorie behandelen waarin deze vergelijking inderdaad mogelijk is.

HOOFDSTUK 2

2.1. MARTINGAALTHEORIE EN DE THEORIE DER STOCHASTISCHE INTEGRALEN

Zij (Ω, F, P) een kansruimte en $\{F_t\}_{t\in T}$ een familie deel σ-algebra's van F, waarbij T een indexverzameling van reële getallen voorstelt ($T = \mathbb{N}, [0,1], [0,\infty)$).
Stel $\{F_t\}$ is *toenemend* dwz. $F_s \subset F_t$ als $s < t$ en $\{F_t\}$ is *rechtskontinu* dwz. $F_t = \cap_{h>0} F_{t+h}$.

We eisen verder dat de σ-algebra's $\{F_t\}$ de volgende eigenschap hebben: Als $A \in F$, $P(A) = 0$ en $B \subset A$ dan volgt hieruit dat $B \in F_t$ voor elke $t \in T$. De σ-algebra's $\{F_t\}$ heten *uitgebreid volledig* als aan deze eis is voldaan.

De generalisatie van een rij stochastische grootheden $\{X_n\}_{n\in\mathbb{N}}$ is een collectie stochastische grootheden $\{X_t\}$ geïndiceerd door een parameter $t \in$ interval $I \subset \mathbb{R}_+$. Dit noemen wij een *stochastisch proces*.

DEFINITIE 2.1.1. Een *stochastisch proces* is een collectie $\{X_t\}$ van stochastische grootheden op (Ω, F, P) waarbij t een interval $I \subset \mathbb{R}_+$ doorloopt.

Notaties: $\{X(t)\}$; $\{X_t\}$.

Voor vaste $\omega \in \Omega$ geldt: $X(t,\omega)$ is een functie (*een pad*) $x(t)$ op I. Een bekend voorbeeld van een stochastische proces is het Wiener proces dat later ter sprake zal komen.

We zeggen dat een stochastisch proces $\{X_t\}$ "*hoort bij*" $\{F_t\}$ als $\omega \to X_t(\omega)$, $\omega \in \Omega$ een F_t-meetbare functie is voor elke $t \in I$.

$L^p(\Omega, F, P)$ met $1 \le p < \infty$, is de vektorruimte van stochastische grootheden X waarvoor geldt dat $\int_\Omega |X(\omega)|^p dP(\omega) < \infty$ en $L^\infty(\Omega, F, P)$ is de vektorruimte van stochastische grootheden X waarvoor geldt dat $\sup\{q: P(\{\omega: |X(\omega)| > q\}) > 0\} < \infty$. Als dit laatste geldt dan heet X *essentieel begrensd*.

DEFINITIE 2.1.2. Als $X(t) \in L^1(\Omega,F,P)$ hoort bij F_t voor elke $t \in I$ dan heet het proces $\{X(t)\}$ een

- *martingaal* als $E(X(t)|F_s) = X(s)$ met kans 1 voor alle $s \leq t$
- *submartingaal* als $E(X(t)|F_s) \geq X(s)$ met kans 1 voor alle $s \leq t$
- *supermartingaal* als $E(X(t)|F_s) \leq X(s)$ met kans 1 voor alle $s \leq t$.

VOORBEELD 2.1.3. We nemen $T = \mathbb{N}$.
Als $Y_1,Y_2,\ldots$ o.o. stochastische grootheden zijn met $EY_i = 0$ voor $i=1,2,\ldots$ en als $X_n = \Sigma_{i=1}^n Y_i$, met $n = 1,2,\ldots$, dan is de rij $X_1,X_2,\ldots$ een martingaal t.o.v. de σ-algebra's $\{F_n\}$ met $F_n = \sigma(X_1,\ldots,X_n)$. Immers $E(X_n|F_m) =$ $= E(X_n|X_1,\ldots,X_m) = X_m$ voor $1 \leq m \leq n$.
Als $0 < EY_i < \infty$ voor $i = 1,2,\ldots$ dan is de rij $X_1,X_2,\ldots$ een submartingaal en als $-\infty < EY_i < 0$ voor $i = 1,2,\ldots$ dan is de rij $X_1,X_2,\ldots$ een supermartingaal.

Stel dat de afbeelding $t \to M(t,\omega)$ rechtskontinue paden met linkerlimieten heeft voor P-bijna alle ω.
Van nu af werken we met $T = [0,\infty)$ tenzij anders vermeld.

DEFINITIE 2.1.4. De martingaal $\{M(t)\}$ heet *kwadratisch integreerbaar* als $\sup_{t\in[0,\infty)} E(M(t))^2 < \infty$ en $M(0) = 0$ met kans 1.

De ruimte van rechtskontinue kwadratisch integreerbare martingalen noteren we met M^2 en geven we de seminorm $\{E(M(\infty))^2\}^{\frac{1}{2}}$.

OPMERKING 2.1.5. Daar $\{M(t)\}$ kwadratisch integreerbaar is, bestaat $\lim_{t\to\infty} M(t,\omega) = M(\infty,\omega)$ voor P-bijna alle $\omega \in \Omega$. (zie MEYER [19] stelling 6, p.131).

BEWERING 2.1.6. *Als* $\{M(t)\} \in M^2$ *dan is* $\{M^2(t)\}$ *een submartingaal.*

STAVING. Volgens de ongelijkheid van Jensen geldt:

$$E(M^2(t)|F_s) \geq (E(M(t)|F_s))^2 = M^2(s) \qquad \text{voor alle } s \leq t. \qquad \square$$

Als $\{M(t)\}$ een submartingaal is, dan is $\{-M(t)\}$ een supermartingaal, immers $E(-M(t)|F_s) = -E(M(t)|F_s) \leq -M(s)$ voor alle $s \leq t$.
Stel dat $0 = t_0 < t_1 < t_2 < \ldots < t_n < \infty$ een partitie van het interval $[0,\infty)$ is en $\{M(t)\} \in M^2$.

DEFINITIE 2.1.7. Het proces $\{H(t)\}$ heet een *eenvoudig proces* als $H(t)$ aan de volgende eisen voldoet:

$H(0) = 0$ met kans 1,

op $(0,t_1]$ is $H(t) = K_0$ waarbij K_0 een F_0-meetbare begrensde stochastische grootheid is,

op $(t_1,t_2]$ is $H(t) = K_1$ waarbij K_1 een F_{t_1}-meetbare begrensde stochastische grootheid is, ...,

op $(t_{n-1},t_n]$ is $H(t) = K_{n-1}$ waarbij K_{n-1} een $F_{t_{n-1}}$-meetbare begrensde stochastische grootheid is. Na t_n geldt $H(t) = 0$.

We definiëren voor eenvoudige processen $\{H(t)\}$ en kwadratisch integreerbare martingalen $\{M(t)\}$: $\int_0^\infty H(s)\,dM(s) = \Sigma_{i=1}^n H(t_i)\{M(t_i)-M(t_{i-1})\}$ en tevens $\int_0^t H(s)\,dM(s) = \int_0^\infty H(s)\,I_{[0,t]}(s)\,dM(s)$. Merk op dat $H(s)I_{[0,t]}(s)$ een eenvoudig proces is op $[0,\infty)$ met extra partitie punt t. We noteren de stochastische grootheid $\int_0^t H(s)\,dM(s)$ met $H{\circ}M(t)$.

OPMERKING 2.1.8. $H{\circ}M(t)$ noemen we de *stochastische integraal* gedefinieerd voor eenvoudige processen.

EIGENSCHAP 2.1.9. $\{H{\circ}M(t)\} \in M^2$, want $E(\int_0^t H(s)dM(s)\,|\,F_u) =$

$$E\left(\sum_{i=1}^{k-1} H(t_i)\{M(t_i)-M(t_{i-1})\} + H(t)\{M(t)-M(t_{k-1})\}\,\Big|\,F_u\right)$$

$$= \sum_{i=1}^{m} H(t_i)\{M(t_i)-M(t_{i-1})\} + H(u)\{M(u)-M(t_m)\} \qquad \text{als } t \in (t_{k-1},t_k]$$

en $u \in (t_m,t_{m+1}]$ met $u \le t$.

Dus $E(\int_0^t H(s)dM(s)\,|\,F_u) = \int_0^u H(s)dM(s)$ voor alle $u \le t$.

$$\sup_t E\left(\int_0^t H(s)dM(s)\right)^2 = E\left(\int_0^\infty H(s)dM(s)\right)^2 \qquad \text{omdat } \int_0^t H(s)dM(s) \text{ een martingaal is}$$

$$= E\left(\sum_{i=1}^n H(t_i)\,(M(t_i)-M(t_{i-1}))\right)^2 = \sum_{i=1}^n E\Big(H(t_i)\,(M(t_i)-M(t_{i-1}))\Big)^2$$

$$< \infty$$

omdat $H(t_i)$ begrensd is en $M(t) \in M^2$.

DEFINITIE 2.1.10. Zij $\{B(t)\}$ een stochastisch proces horend bij de familie σ-algebra's $\{F_t\}$. $\{B(t)\}$ heet een *groeiend proces* als

(i) Voor P-bijna alle $\omega \in \Omega$, $B(0,\omega) = 0$ en de functie $t \to B(t,\omega)$ is niet-dalend en rechtskontinu.

(ii) $EB(t) < \infty$ voor alle $t \in [0,\infty)$.

Een stochastisch proces $\{X(t)\}$ heet *linkskontinu* als de paden $t \to X(t,\omega)$ voor P-bijna alle ω linkskontinu zijn.

DEFINITIE 2.1.11. Een stochastisch proces $\{X(t)\}$ heet *voorspelbaar* als $X(t,\omega)$ meetbaar is t.o.v. de σ-algebra op $\mathbb{R}_+ \times \Omega$ voortgebracht door de eenvoudige processen.

DEFINITIE 2.1.12. Een stochastisch proces $\{Y(t)\}$ heet *niet-negatief* als met kans 1 alle paden niet-negatief zijn.

OPMERKING 2.1.13. Een stochastisch proces $\{A(t)\}$ heet een *voorspelbaar groeiend* proces als $\{A(t)\}$ zowel aan definitie 2.1.10. als aan definitie 2.1.11. voldoet.

OPMERKING 2.1.14. In de literatuur wordt in plaats van het begrip voorspelbaar groeiend ook het begrip natuurlijk groeiend proces gebruikt. Daar deze begrippen equivalent zijn (zie DOLÉANS [10] p. 874-876) gebruiken wij alleen de eerst genoemde term.

STELLING 2.1.15.

De ontbindingsstelling van Doob-Meyer: *Stel* $\{X(t)\}$ *is een rechtskontinue niet-negatieve submartingaal. Dan bestaat er een uniek voorspelbaar groeiend proces* $\{A(t)\}$ *zodat* $\{X(t)-A(t)\}$ *een martingaal is (uniek betekent: afgezien van een P-nulverzameling uniek bepaald).*

BEWIJS. zie MEYER [19], hoofdstuk VI en VII.

Stel $\{M(t)\} \in \mathcal{M}^2$ en $\{A(t)\}$ is het unieke voorspelbare groeiende proces korresponderend met $\{M^2(t)\}$.

Definieer $\langle M,M\rangle(t) = A(t)$ zodat $\{M^2(t) - \langle M,M\rangle(t)\}$ een martingaal is. Zij voorts $\{M_1(t)\}$ en $\{M_2(t)\}$ in $\mathcal{M}^2$.

Definieer $\langle M_1,M_2\rangle(t) = \frac{1}{4}[\langle M_1+M_2,M_1+M_2\rangle(t) - \langle M_1-M_2,M_1-M_2\rangle(t)]$.

OPMERKING 2.1.16. Men kan $\langle M_1,M_2\rangle(t)$ opvatten als een soort inwendig produkt van kwadratisch integreerbare martingalen.

Eigenschappen van $\{\langle M_1,M_2\rangle(t)\}$ 2.1.17.

(a) $\{\langle M_1,M_2\rangle(t) - M_1(t)M_2(t)\}$ is een martingaal

want $\{(M_1(t)-M_2(t))^2 - \langle M_1-M_2,M_1-M_2\rangle(t)\}$ is een martingaal

$\underline{\{(M_1(t)+M_2(t))^2 - \langle M_1+M_2,M_1+M_2\rangle(t)\}}$ is een martingaal $-$

$\{-4M_1(t)M_2(t) + \langle M_1+M_2,M_1+M_2\rangle(t) - \langle M_1-M_2,M_1-M_2\rangle(t)\}$ is een martingaal

dus $\{-M_1(t)M_2(t) + \langle M_1,M_2\rangle(t)\}$ is een martingaal.

(b) $\{\langle M_1,M_2\rangle(t)\}$ is het verschil van twee voorspelbare groeiende processen.

(c) $\{\langle M_1,M_2\rangle(t)\}$ is het unieke proces dat voldoet van (a) en (b).

BEWIJS. Analoog aan het bewijs van stelling 21 p.150-151 uit MEYER [19].

(d) $\langle M_1,M\rangle(t) + \langle M_2,M\rangle(t) = \langle M_1+M_2,M\rangle(t)$. Door het uitschrijven van de definitie van $\langle M_1,M_2\rangle(t)$ is bovenstaande triviaal.

DEFINITIE 2.1.18. Als $\{M_1(t)\}$ en $\{M_2(t)\}$ in M^2 dan heten $\{M_1(t)\}$ en $\{M_2(t)\}$ *orthogonaal* als $\langle M_1,M_2\rangle(t) = 0$ voor alle t met kans 1.

LEMMA 2.1.19. *Stel* $\{M_1(t)\}$ en $\{M_2(t)\} \in M^2$. Dan zijn $\{M_1(t)\}$, $\{M_2(t)\}$ *orthogonaal dan en slechts dan als* $\{M_1(t)M_2(t)\}$ *een martingaal is.*

BEWIJS. Eigenschap 2.1.17.(a) impliceert dat als $\langle M_1,M_2\rangle(t) = 0$ voor alle t met kans 1 dat $\{M_1(t)M_2(t)\}$ een martingaal is.
Als $\{M_1(t)M_2(t)\}$ een martingaal is dan zijn $\{M_1(t)\}$, $\{M_2(t)\}$ orthogonaal vanwege eigenschap 2.1.17.(c).

LEMMA 2.1.20. *Als* $\{H(t)\}$ *een eenvoudig proces is en* $\{M(t)\} \in M^2$ *dan geldt dat* $E(\int_0^t H(s)dM(s))^2 = E(\int_0^t H^2(s)d\langle M,M\rangle(s))$.

BEWIJS. Zij $t \in (0,t_1]$.

$EH^2(t)M^2(t) = EH^2(t)\langle M,M\rangle(t)$ immers $E\{H^2(t)(M^2(t) - \langle M,M\rangle(t))\} =$

$E\{H^2(t)\, E(M^2(t)-\langle M,M\rangle(t) \mid F_0)\} = 0.$

Stel $t \in (t_1,t_2]$.

$$E(\int_0^t H(s)dM(s))^2$$

$$= E(H(t_1)(M(t_1)-M(0)) + H(t)(M(t)-M(t_1)))^2$$

$$= E(H^2(t_1)(M(t_1)-M(0))^2 + H^2(t)(M(t)-M(t_1))^2 + 2(H(t_1)H(t)(M(t_1)-M(0))(M(t)-M(t_1))))$$

$$= E(H^2(t_1)E((M(t_1)-M(0))^2|F_0)) + E(H^2(t)E((M(t)-M(t_1))^2|F_{t_1})) +$$

$$2E(H(t_1)(M(t_1)-M(0))\ H(t)\ \underbrace{E(M(t)-M(t_1))|F_{t_1})}_{=0})$$

$$= E(H^2(t_1)E(\langle M,M\rangle(t_1)-\langle M,M\rangle(0)|F_0)) + E(H^2(t)E(\langle M,M\rangle(t)-\langle M,M\rangle(t_1)|F_{t_1})$$

$$= E(H^2(t_1)(\langle M,M\rangle(t_1)-\langle M,M\rangle(0)) + H^2(t)(\langle M,M\rangle(t)-\langle M,M\rangle(t_1)))$$

want $E(M^2(t)-M^2(s)|F_s) = E((M(t)-M(s))^2|F_s) = E(A(t)-A(s)|F_s)$ met

$A(t) = \langle M,M\rangle(t)$ en $s \le t$.

Dit procedé herhalen voor $t \in (t_i,t_{i-1}]$ $i = 2,3....,n$. □

DEFINITIE 2.1.21. Twee processen $\{X(t)\}$ en $\{Y(t)\}$ waarvoor geldt dat $X(t) = Y(t)$ voor elke $t \in [0,\infty)$ met kans 1 heten *ononderscheidbaar*.

STELLING 2.1.22. *Zij* $\{M(t)\} \in M^2$ *en* $\{H(t)\}$ *een eenvoudig proces. Dan geldt* $E(\int_0^\infty |H(s)|\ |d\langle M,N\rangle(s)|) < \infty$ *voor elke* $\{N(t)\} \in M^2$ *en* $\{H\circ M(t)\}$ *is het uniek bepaalde (afgezien van ononderscheidbare versies) element van* M^2 *zodat voor elke* $\{N(t)\} \in M^2$ *en elke* $t \in [0,\infty)$ $\langle H\circ M,N\rangle(t) = \int_0^t H(s)d\langle M,N\rangle(s)$ *met kans* 1.

OPMERKING 2.1.23. Een eindige reële σ-additieve verzamelingsfunctie op 'n σ-algebra B noemen wij een *getekende maat*. Voor zo'n getekende maat ν bestaan er minimale eindige maten ν^+ en ν^- zodanig dat op de σ-algebra geldt $\nu = \nu^+ - \nu^-$ (Jordan decompositie, zie ZAANEN [28] p.318-320). In ons geval is $\nu([0,t]) = \langle M,N\rangle(t)$; $|d\nu| \overset{def}{=} d\nu^+ + d\nu^-$ op de Borel σ-algebra op $[0,\infty)$ voor iedere ω.

<u>BEWIJS</u>. Het proces $\{H(t)\}$ definiëren we door

$$H(t) = \sum_{i=1}^{k-1} I_{(t_i,t_{i+1}]}(t)\,\bar{H}_i + I_{(t_k,\infty)}(t)\,\bar{H}_k \qquad \text{waarbij } 0 = t_1 < t_2 < \ldots < t_k$$

en $\bar{H}_i$ een $\mathcal{F}_{t_i}$-meetbare begrensde stochastische grootheid is, $i = 1,\ldots,k$. Hieruit volgt dat $E(\int_0^\infty |H(s)|\ |d\langle M,N\rangle(s)|) < \infty$ en dmv uitschrijven dat $\langle H\circ M,N\rangle(t) = \int_0^t H(s)d\langle M,N\rangle(s)$ met kans 1.
Er is (afgezien van ononderscheidbare versies) voor elk eenvoudig proces $\{H(t)\}$ maar één element $\{H\circ M(t)\}$ dat bovenstaande eigenschap bezit. Stel dat $\{L(t)\},\{L'(t)\} \in M^2$ zodanig dat $\langle L,N\rangle(t) = \langle L',N\rangle(t) = \int_0^t H(s)d\langle M,N\rangle(s)$ voor elke $t \in [0,\infty)$ met kans 1 en elke $\{N(t)\} \in M^2$. Kies $\{N(t)\} = \{L(t)-L'(t)\}$ dan geldt dat $\langle L-L',L-L'\rangle(t) = 0$. Dus $(L-L')^2(t)$ is een niet-negatieve martingaal (Lemma 2.1.19.) zodat $(L(t)-L'(t))^2 = 0$ voor elke $t \in [0,\infty)$ met kans 1. □

De voorgaande theorie gaan we nu uitbreiden van eenvoudige processen tot voorspelbare processen.

Stel $\{M(t)\} \in M^2$. Definieer $L^2(M)$ als de klasse van alle voorspelbare processen $\{H(t)\}$ waarvoor geldt dat

$$E\left(\int_0^\infty H^2(s)\ d\langle M,M\rangle(s)\right) < \infty.$$

We voorzien $L^2(M)$ van de seminorm

$$\left\{E\left(\int_0^\infty H^2(s)\ d\langle M,M\rangle(s)\right)\right\}^{\frac{1}{2}}.$$

Een proces waarvan alle paden linkskontinu zijn, is voorspelbaar. Een Borel-meetbare functie is voorspelbaar. Voor linkskontinue processen bestaan er ononderscheidbare versies die voorspelbaar zijn. Zij E de deelruimte van $L^2(M)$ bestaande uit de eenvoudige processen. Als men van nu af ononderscheidbare processen zowel in M^2 als in $L^2(M)$ met elkaar identificeert dan kan men bewijzen dat E dicht ligt in $L^2(M)$ in de zin van de norm op $L^2(M)$ en dan geldt:

<u>STELLING 2.1.24</u>. *Stel* $\{M(t)\} \in M^2$. *De afbeelding* $H(t) \to \int_0^t H(s)\ dM(s)$ *van de eenvoudige processen naar de kwadratisch integreerbare martingalen kan uniek worden uitgebreid tot een normbehoudende afbeelding van* $L^2(M)$ *naar* M^2.

<u>OPMERKING 2.1.25</u>. De seminormen $\{E(M(\infty))^2\}^{\frac{1}{2}}$ en $\{E(\int_0^\infty H^2(s)d\langle M,M\rangle(s))\}^{\frac{1}{2}}$ in M^2 respectievelijk $L^2(M)$ zijn normen als ononderscheidbare versies met elkaar

geïdentificeerd worden.

Het bewijs van stelling 2.1.24. volgt uit de volgende twee lemma's.

LEMMA 2.1.26. *M^2 is een Banach ruimte.*

BEWIJS. Stel dat $\{M_n(t)\}$ een fundamentaalrij is, dwz. $E(M_n(\infty)-M_m(\infty))^2 \to 0$ als $n,m \to \infty$.

$L^2(\Omega,F,P)$ is een Banach ruimte als P-bijna overal gelijke stochastische variabelen geïdentificeerd worden. Er bestaat dus een stochast $M(\infty) \in L^2(\Omega,F,P)$ zodanig dat $E(M_n(\infty)-M(\infty))^2 \to 0$ als $n \to \infty$.

Definieer $M(t) = E(M(\infty)|F_t)$.

Dus de rij kwadratisch integreerbare martingalen konvergeert naar de limiet $M(t)$ in de norm van M^2.

$\{M(t)\} \in M^2$.

Andere eigenschappen van Banach ruimten zijn triviaal. □

LEMMA 2.1.17. *Stel* V *is een genormeerde vektorruimte,* U *is een Banach ruimte,* W *is een deelruimte van* V *die dicht in* V *ligt en* f *is een kontinue lineaire afbeelding van* W *in* U. *Dan bestaat er een unieke kontinue lineaire afbeelding* $\bar{f}$ *van* V *in* U *waarvan de restrictie tot* W f *is.*

BEWIJS. Zie DIEUDONNÉ [9] p.103-105.

STELLING 2.1.28. *Zij* $\{M(t)\} \in M^2$ *en* $\{H(t)\} \in L^2(M)$. *Dan geldt* $E(\int_0^\infty H(s)|\,|d\langle M,N\rangle(s)|) < \infty$ voor elke $\{N(t)\} \in M^2$ *en er bestaat een uniek bepaald (afgezien van ononderscheidbare versies) element van* M^2 *genoteerd door* $H\circ M$ *zodat voor elke* $\{N(t)\} \in M^2$ *en elke* $t \in [0,\infty)$ *geldt* $\langle H\circ M,N\rangle(t) = \int_0^t H(s)d\langle M,N\rangle(s)$ *met kans* 1.

BEWIJS. (schets)

Stel dat de processen $\{H(t)\}$ en $\{K(t)\}$ gedefinieerd zijn door

$$H(t) = \sum_{i=1}^{k-1} I_{(t_i,t_{i+1}]}(t)\bar{H}_i + I_{(t_k,\infty)}(t)\bar{H}_k.$$

$$K(t) = \sum_{i=1}^{k-1} I_{(t_i,t_{i+1}]}(t)\bar{K}_i + I_{(t_k,\infty)}(t)\bar{K}_k \quad \text{waarbij } 0=t_1<t_2<\ldots<t_k$$

en $\bar{H}_i$ en $\bar{K}_i$ F_{t_i}-meetbare begrensde stochastische variabelen zijn. Dan geldt (met $t_{k+1}=\infty$)

$$E\left(\int_0^\infty H(t)K(t)d\langle M,N\rangle(t)\right) = E\left(\sum_{i=1}^k \overline{H}_i\overline{K}_i E\left(\langle M,N\rangle(t_{i+1})-\langle M,N\rangle(t_i)\,|\,F_{t_i}\right)\right)$$

$$= E\left\{\sum_{i=1}^k \overline{H}_i\overline{K}_i\ E\left(M(t_{i+1})N(t_{i+1})-M(t_i)N(t_i)\,|\,F_{t_i}\right)\right\}$$

$$= E\left(\sum_{i=1}^k \overline{H}_i\overline{K}_i\ (M(t_{i+1})-M(t_i))(N(t_{i+1})-N(t_i))\right)$$

$$\leq E\left(\left\{\sum_{i=1}^k \overline{H}_i^{\,2}(M(t_{i+1})-M(t_i))^2\right\}^{\frac{1}{2}}\left\{\sum_{i=1}^k \overline{K}_i^{\,2}(N(t_{i+1})-N(t_i))^2\right\}^{\frac{1}{2}}\right)$$

$$\leq \left\{E\left(\sum_{i=1}^k \overline{H}_i^{\,2}(\langle M,M\rangle(t_{i+1})-\langle M,M\rangle(t_i))\right)\right\}^{\frac{1}{2}}\left\{E\left(\sum_{i=1}^k \overline{K}_i^{\,2}(\langle N,N\rangle(t_{i+1})-\langle N,N\rangle(t_i))\right)\right\}^{\frac{1}{2}}.$$

Door limietovergang volgt uit definitie 2.1.11. dat de ongelijkheid $E(\int_0^\infty H(t)K(t)d\langle M,N\rangle(t)) \leq (E\int_0^\infty H^2(t)d\langle M,M\rangle(t))^{\frac{1}{2}}\ (E\int_0^\infty K^2(t)d\langle N,N\rangle(t))^{\frac{1}{2}}$ ook geldt als $\{H(t)\}$ en $\{K(t)\}$ voorspelbare processen zijn.

Stel μ is een getekende maat op de σ-algebra S op $\mathbb{R}_+ \times \Omega$, voortgebracht door de voorspelbare processen, gedefinieerd door

$$\mu(A) = \int_\Omega \left\{\int_0^\infty I_A(s,\omega)d\langle M,N\rangle(s)\right\} dP(\omega) \qquad \text{voor } A \in S.$$

Op dezelfde manier definiëren we de maat $|\mu|$ op S door

$$|\mu|(A) = \int_\Omega \left\{\int_0^\infty I_A(s,\omega)\ |d\langle M,N\rangle(s)|\right\} dP(\omega).$$

Het is duidelijk dat μ $|\mu|$-absoluut kontinu is.

Stel L is een Radon-Nikodym afgeleide $\frac{d\mu}{d|\mu|}$ op S.

Dan is L S-meetbaar dus voorspelbaar.

L kan zo gekozen worden dat $L(s,\omega) = \pm 1$ (toepassing van Hahn decompositie van $\mathbb{R}_+ \times \Omega$ in verzamelingen $A,B \in S$ levert $\mathbb{R}_+ \times \Omega = A \cup B$, $A \cap B = \emptyset$ en $\mu^+(C) = \mu(C\cap A) \geq 0$, $\mu^-(C) = -\mu(C\cap B) \geq 0$, met $C \in S$.
Kies $L = I_A - I_B$; zie ZAANEN [28] p.320-321).
Dus $Ld\mu = |d\mu| = d|\mu|$.

Verder $E\left(\int_0^t H(s)K(s)\ |d\langle M,N\rangle(s)|\right)$

$$= \int_\Omega \left\{\int_0^\infty I_{[0,t]}(s)\ H(s)K(s)\ |d\langle M,N\rangle(s)|\right\} dP(\omega)$$

$$= \int_{\mathbb{R}_+\times\Omega} I_{[0,t]}(s)\ H(s)K(s)\ d|\mu|$$

$$= \int_{\mathbb{R}_+\times\Omega} I_{[0,t]}(s)\ H(s)K(s)\ L(s)\ d\mu$$

$$= E\left(\int_0^t H(s)K(s)\ L(s)\ d\langle M,N\rangle(s)\right)$$

$$\leq \left(E\int_0^t H^2(s)\ d\langle M,M\rangle(s)\right)^{\frac{1}{2}} \left(E\int_0^t K^2(s)\ d\langle N,N\rangle(s)\right)^{\frac{1}{2}} < \infty.$$

Neem $K(t) \equiv 1$ en $H(t) \geq 0$ met kans 1.
De Lebesgue-Stieltjes integraal $\int_0^t H(s)d\langle M,N\rangle(s)$ is voor elke t nu met kans 1 gedefinieerd.
We gaan nu bewijzen dat $\langle H\circ M,N\rangle(t) = \int_0^t H(s)d\langle M,N\rangle(s)$ met kans 1.

$$E|\langle M,N\rangle(t)| \leq (EM^2(\infty))^{\frac{1}{2}}\ (EN^2(\infty))^{\frac{1}{2}} \quad \text{want}$$

$\langle \alpha M+\beta N,\ \alpha M+\beta N\rangle(t) \geq 0$ met kans 1, dus
als $\alpha = \langle N,N\rangle^{\frac{1}{2}}(t)$ en $\beta = -\langle M,M\rangle^{\frac{1}{2}}(t)$ dan geldt dat

$$2\langle M,M\rangle(t)\langle N,N\rangle(t) \geq 2\langle M,M\rangle^{\frac{1}{2}}(t)\langle N,N\rangle^{\frac{1}{2}}(t)\langle M,N\rangle(t).$$

Volgens de ongelijkheid van Cauchy-Schwarz geldt dat

$$E\langle M,N\rangle(t) \leq E\langle M,M\rangle^{\frac{1}{2}}(t)\langle N,N\rangle^{\frac{1}{2}}(t) \leq (E\langle M,M\rangle(t))^{\frac{1}{2}}\ (E\langle N,N\rangle(t))^{\frac{1}{2}}$$

$$\leq (E\langle M,M\rangle(\infty))^{\frac{1}{2}}(E\langle N,N\rangle(\infty))^{\frac{1}{2}} = (EM^2(\infty))^{\frac{1}{2}}(EN^2(\infty))^{\frac{1}{2}}.$$

Analoog $E(-\langle M,N\rangle(t)) \leq (EM^2(\infty))^{\frac{1}{2}}(EN^2(\infty))^{\frac{1}{2}}$ (neem $\alpha = \langle N,N\rangle^{\frac{1}{2}}(t)$ en $\beta = \langle M,M\rangle^{\frac{1}{2}}(t)$).

Dus als $M_n \to M$ in $\mathcal{M}^2$ dan volgt hieruit dat

$$E|\langle M_n,N\rangle(t) - \langle M,N\rangle(t)| = E|\langle M_n-M,N\rangle(t)|$$

$$\leq (E(M_n-M)^2(\infty))^{\frac{1}{2}}(EN^2(\infty))^{\frac{1}{2}} \to 0.$$

Dus $\langle M_n,N\rangle(t) \to \langle M,N\rangle(t)$ in waarschijnlijkheid, immers

$$P\{|\langle M_n,N\rangle(t) - \langle M,N\rangle(t)| \geq \varepsilon\} \leq \frac{1}{\varepsilon}E|\langle M_n,N\rangle(t)-\langle M,N\rangle(t)| \to 0 \text{ als } n\to\infty.$$

Verder als $H_n \to H$ in $L^2(M)$, met $\{H_n(t)\}$ een eenvoudig proces voor alle n, dan

$$E\left|\int_0^t H_n(s)d\langle M,N\rangle(s) - \int_0^t H(s)d\langle M,N\rangle(s)\right|$$

$$\leq E\int_0^t |H_n(s)-H(s)|\ |d\langle M,N\rangle(s)| \leq E\int_0^\infty |(H_n(s)-H(s)|\ |d\langle M,N\rangle(s)|$$

$$\leq \left(E\int_0^\infty (H_n(s)-H(s))^2 d\langle M,M\rangle(s)\right)^{\frac{1}{2}}(E\langle N,N\rangle(\infty))^{\frac{1}{2}}$$

Dus $\int_0^t H_n(s)d\langle M,N\rangle(s) \to \int_0^t H(s)d\langle M,N\rangle(s)$ in waarschijnlijkheid.

Er is een deelrij (n_k) zodat $\int_0^t H_{n_k}(s)d\langle M,N\rangle(s) \to \int_0^t H(s)d\langle M,N\rangle(s)$ met kans 1 (zie TUCKER [26] stelling 4.4.2. p.101).

Er is een verdere deelrij (n_l') van (n_k) zodat

$$\langle H_{n_l'}\circ M,N\rangle(t) \to \langle H\circ M,N\rangle(t) \text{ met kans 1.}$$

Dus $$\langle H\circ M,N\rangle(t) = \int_0^t H(s)d\langle M,N\rangle(s) \qquad \text{voor alle } t, \text{ met kans 1.}$$

De uniciteit van $H\circ M(t)$ kan analoog aan het bewijs van Stelling 2.1.22. bewezen worden. □

DEFINITIE 2.1.29. Als $\{M(t)\} \in M^2$ en $\{H(t)\} \in L^2(M)$ dan wordt de *stochastische integraal* $\int_0^t H(s)dM(s)$ gedefinieerd door $H\circ M(t)$ voor elke t met kans 1.

OPMERKING 2.1.30. Voor $\{H(t)\} \in E$ hebben we bewezen dat $\{H\circ M(t)\} \in M^2$ en $E(\int_0^t H(s)dM(s))^2 = E(\int_0^t H^2(s)d\langle M,M\rangle(s))$. De deelruimte E van $L^2(M)$ ligt dicht in $L^2(M)$. Door limietovergang gelden bovenstaande eigenschappen ook voor $\{H(t)\} \in L^2(M)$.

OPMERKING 2.1.31. Een stochastische integraal behoeft geen Lebesgue-Stieltjes integraal te zijn omdat de functie $s \to M(s,\omega)$, $s \in [0,t]$ niet noodzakelijk van begrensde variatie is. Sterker nog:
Een kontinue martingaal is van begrensde variatie dan en slechts dan als hij met kans 1 konstant is (zie DELLACHERIE [8] p.111). In de volgende paragraaf zullen we het verband tussen een stochastische integraal en een Lebesgue-Stieltjes integraal geven.

STELLING 2.1.32. *Stel* $\{M(t)\}, \{N(t)\} \in M^2$, $\{H(t)\} \in L^2(M)$ *en* $\{K(t)\} \in L^2(N)$.

Dan geldt $E\left(\int_0^\infty |H(s)K(s)|\ |d\langle M,N\rangle(s)|\right) < \infty$ *en*

$$\langle H\circ M, K\circ N\rangle(t) = \int_0^t H(s)K(s)d\langle M,N\rangle(s) \text{ met kans 1 voor alle } t \in [0,\infty).$$

BEWIJS. (schets)
Het eerste gedeelte van het bewijs is reeds bewezen in Stelling 2.1.28.
Kies nl. $H(t) = |H(t)|$ en $K(t) = |K(t)|$ voor alle t met kans 1 dan

$$E\left(\int_0^\infty |H(s)K(s)|\ |d\langle M,N\rangle(s)|\right) < \infty.$$

Per definitie is $\int_0^t H(s)d\langle M,N\rangle(s) = H\circ\langle M,N\rangle(t)$.
Dus $\langle H\circ M, K\circ N\rangle(t) = H\circ\langle M,K\circ N\rangle(t) = H\circ\langle K\circ N,M\rangle(t) = H\circ K\circ\langle M,N\rangle(t)$.
Nu geldt dat

$$\int_0^t H(s)d\left(\int_0^s K(u)d\langle M,N\rangle(u)\right) = \int_0^t H(s)K(s)d\langle M,N\rangle(s) \text{ voor elke } t \in [0,\infty)$$

met kans 1 omdat de relatie waar is voor trapfuncties H(t).

Dus $<H\circ M,\ K\circ N>(t) = H\circ K\circ <M,N>(t) = HK\circ <M,N>(t)$. □

GEVOLGEN 2.1.33.

(i) Als $\{M(t)\} \in \mathcal{M}^2$ en $\{H(t)\} \in \mathcal{L}^2(M)$ dan geldt dat

$$<H\circ M,\ H\circ M>(t) = <\int_0^t H(s)dM(s),\ \int_0^t H(s)dM(s)>$$

$$= \int_0^t H^2(s)d<M,M>(s) \quad \text{voor alle } t \in [0,\infty) \text{ met kans 1.}$$

(ii) Als $\{M_1(t)\}$, $\{M_2(t)\} \in \mathcal{M}^2$ en $\{H_1(t)\} \in \mathcal{L}^2(M_1)$ en $\{H_2(t)\} \in \mathcal{L}^2(M_2)$ dan als $\{M_1(t)\}$, $\{M_2(t)\}$ orthogonaal zijn, geldt dat $\{\int_0^t H_1(s)dM_1(s)\}$, $\{\int_0^t H_2(s)dM_2(s)\}$ ook orthogonaal zijn.

2.2. TELPROCESSEN

Zij $(\Omega,\mathcal{F},P)$ een kansruimte.

DEFINITIE 2.2.1. Een meerdimensionaal stochastisch proces $\{\vec{N}(t)\} = \{(N_1(t),N_2(t),\ldots,N_k(t));\ t \in [0,\infty)\}$ heet een *meerdimensionaal telproces* als de paden $\vec{N}(t,\omega) = (N_1(t,\omega),N_2(t,\omega),\ldots,N_k(t,\omega))$ voor P-bijna alle $\omega \in \Omega$ de volgende eigenschappen bezitten:

(a) $N_i(t,\omega)$ is voor $i = 1,2,\ldots,k$ een rechtskontinue niet-dalende trapfunctie. Deze trapfunctie maakt eindig veel sprongen van hoogte 1. Voor elke i geldt dat $N_i(0,\omega) = 0$.

(b) Als $N_i(t,\omega) \neq N_i(t-,\omega)$ dan geldt dat $N_j(t,\omega) = N_j(t-,\omega)$ voor alle $j \neq i;\ i = 1,2,\ldots,k$.

$\mathcal{N}_t$ is per definitie de σ-algebra voortgebracht door $\{\vec{N}(s);\ s\le t\}$. De familie σ-algebra's $\{\mathcal{N}_t;\ t\in[0,\infty)\}$ is rechtskontinu dwz. $\bigcap_{h>0}\mathcal{N}_{t+h} = \mathcal{N}_t$ voor elke $t \in [0,\infty)$ en is toenemend dwz. $\mathcal{N}_s \subset \mathcal{N}_t\ \forall s \le t$.
$\{\mathcal{F}_t:\ t\in[0,\infty)\}$ is een andere rechtskontinue familie deel σ-algebra's van $\mathcal{F}$ zodanig dat $\mathcal{N}_t \subset \mathcal{F}_t$ voor elke $t \in [0,\infty)$.

We eisen dat $\mathcal{F}_t$ en $\mathcal{N}_t$ voor elke t uitgebreid volledig zijn met betrekking tot de kansmaat P.

Het is vaak mogelijk om een telproces te beschrijven door middel van een intensiteitsproces.

DEFINITIE 2.2.2. Een *intensiteitsproces* is een niet-negatief proces $\{\vec{\Lambda}(t)\} = \{(\Lambda_1(t),\Lambda_2(t),\ldots,\Lambda_k(t))\}$ horend bij $\{\mathcal{F}_t\}$ zodanig dat

(a) De paden van $\Lambda_i(t)$ voor $i = 1,2,\ldots,k$ linkskontinu zijn met rechterlimieten in elk punt.

(b) $M_i(t) = N_i(t) - \int_0^t \Lambda_i(s)ds$ is voor $i = 1,2,\ldots,k$ een martingaal t.o.v. $\{\mathcal{F}_t\}$.

OPMERKING 2.2.3. Eis (a) van Definitie 2.2.2. betekent dat voor $i=1,2,\ldots,k$ $\Lambda_i(t)$ een voorspelbaar proces is.

STELLING 2.2.4. *Als* $\sup_{t\in[0,\infty)} E(N_i(t)) < \infty$ *voor* $i = 1,2,\ldots,k$ *dan is voor* $i = 1,2,\ldots,k$ $\{M_i(t)\}$ *een kwadratisch integreerbare martingaal t.o.v.* $\{\mathcal{F}_t\}$

waarvoor geldt dat (1) $\langle M_i, M_j\rangle(t) = 0$ *voor* $i \neq j$

(2) $\langle M_i, M_i\rangle(t) = \int_0^t \Lambda_i(s)ds.$

BEWIJS. Zie BOEL, VARAIYA & WONG [5] p.1005-1007.

OPMERKING 2.2.5. Als het intensiteitsproces $\{\Lambda(t)\}$ bestaat, dan is het uniek (afgezien van ononderscheidbare versies).

VOORBEELD 2.2.6. Stel dat $\{N(t)\}$ een Poisson proces voorstelt (zie definitie 2.2.16) met parameter λ. Zij N_t de σ-algebra voortgebracht door $N(s)$ voor alle $s \leq t$. Dan is $\Lambda(t) = \lambda$. Immers

$$E\left(N(t) - \int_0^t \lambda ds \middle| N_s\right) = E(N(t)-\lambda t|N_s)$$

$$= E(N(t)-N(s)-(\lambda t-\lambda s)+N(s)-\lambda s|N_s)$$

$$= \lambda(t-s) - (\lambda t-\lambda s) + N(s) - \lambda s = N(s) - \lambda s.$$

Dus $M(t) = N(t) - \lambda t$ is een martingaal t.o.v. de σ-algebra's $\{N_t\}$.
Verder is $E(M^2(t)-M^2(s)-\lambda(t-s)|N_s) = 0$.

OPMERKING 2.2.7. In de literatuur wordt ook vaak als definitie van een intensiteitsproces gegeven:
Een intensiteitsproces $\{\vec{U}(t)\}$ is een stochastisch proces dat voor alle t voldoet aan (a) en/of (b).

(a) $$U_i(t) = \lim_{h\downarrow 0} \frac{1}{h} E(N_i(t+h)-N_i(t)|F_t) \qquad \text{voor } i = 1,2,\ldots,k.$$

(b) $$U_i(t) = \lim_{h\downarrow 0} \frac{1}{h} P(N_i(t+h)-N_i(t)=1|F_t) = \lim_{h\downarrow 0} \frac{1}{h}\left[1-P(N_i(t+h)-N_i(t)=0|F_t)\right]$$

voor $i = 1,2,\ldots,k.$

Als de komponentprocessen van $\{\vec{U}(t)\}$ zowel aan (a) als (b) voldoen, dan gelden de volgende twee eigenschappen.

(1) $$\lim_{h\downarrow 0} \frac{1}{h} E\left[(N_i(t+h) - N_i(t))\, I_{\{N_i(t+h)-N_i(t)>1\}} \middle| F_t\right] = 0$$

(2) $$\lim_{h\downarrow 0} \frac{1}{h} P\left[N_i(t+h) - N_i(t) > 1 \middle| F_t\right] = 0.$$

BEWIJS. Zie appendix.

Definieer $\{\vec{\Lambda}^*(t)\}$ als de rechtskontinue modificatie van $\{\vec{\Lambda}(t)\}$ (dwz. $\vec{\Lambda}^*(t,\omega) = \Lambda(t+,\omega)$ voor alle $\omega \in \Omega$ en alle $t \in [0,\infty)$).

STELLING 2.2.8. *Als* $\{\Lambda_i(t)\}$ *voor* $i = 1,2,\ldots,k$ *naar boven begrensd wordt door één niet-negatieve integreerbare stochastische grootheid, dan geldt dat* $\{\Lambda_i^*(t)\}$ *voor* $i = 1,2,\ldots,k$ *voldoet aan de eisen* (a) *en* (b) *van Opmerking* 2.2.7.

BEWIJS. Zie appendix.

OPMERKING 2.2.9. Als $\bar{N}(t) = \Sigma_{i=1}^k N_i(t)$ en $\bar{\Lambda}(t) = \Sigma_{i=1}^k \Lambda_i(t)$ dan is $\bar{N}(t) - \int_0^t \bar{\Lambda}(s)ds$ een martingaal. Als $\{\bar{\Lambda}(t)\}$ naar boven begrensd wordt door één niet-negatieve integreerbare stochastische grootheid, dan voldoet $\{\bar{\Lambda}^*(t)\}$ aan Stelling 2.2.8. en dan geldt dat $\lim_{h\downarrow 0} \frac{1}{h} P(\bar{N}(t+h)-\bar{N}(t)>1 \mid F_t) = 0$, dwz. de kans dat twee of meer komponentprocessen veranderen van waarde in het interval $(t,t+h]$, gegeven F_t, is gelijk aan $o(h)$.
Verder geldt dat $\lim_{h\downarrow 0} \frac{1}{h} E((\bar{N}(t+h)-\bar{N}(t)) I_{\{\bar{N}(t+h)-\bar{N}(t)>1\}} \mid F_t) = 0$.

DEFINITIE 2.2.10.

Een F-meetbare functie $T\colon \Omega \to [0,\infty]$ heet een *stoptijd* t.o.v. de familie σ-algebra's $\{F_t\}_{t\in[0,\infty)}$ als $\{\omega\colon T(\omega)\le t\} \in F_t$ voor alle $t \in [0,\infty)$.
Zij F_∞ de σ-algebra voortgebracht door de collectie $\bigcup_{t\in[0,\infty)} F_t$, dan heet de σ-algebra van eventualiteiten $B \in F_\infty$, die voldoen aan $B \cap \{\omega\colon T(\omega)\le t\} \in F_t$ voor elke $t \in [0,\infty)$
de σ-algebra van eventualiteiten voorafgaand aan T. Deze σ-algebra wordt aangeduid met F_T.

VOORBEELD 2.2.11. Zij $A = (a,b)$ een open interval in $\mathbb{R}$ en zij $\{X(t)\}$ een stochastisch proces met rechtskontinue paden horend bij $\{F_t\}$. Dan is de functie $T\colon \Omega \to [0,\infty]$ gedefinieerd door

$$T(\omega) = \begin{cases} \inf\{s\colon s\in[0,\infty);\ X(s,\omega)\in A\} & \text{als } \{s\in[0,\infty)\colon X(s,\omega)\in A\} \neq \emptyset \\ \infty & \text{anders} \end{cases}$$

een stoptijd t.o.v. $\{F_t\}$. Deze stoptijd heet de tijd van binnenkomst in A.

T is een stoptijd omdat

(i) $\{\omega: T(\omega)<t\} = \bigcup_{\substack{s<t \\ s \text{ rationaal}}} \{\omega\in\Omega: X_s(\omega)\in A\} \in F_t$ voor elke $t \in [0,\infty)$

waarbij de gelijkheid volgt uit de rechtskontinuïteit van de functies $s \to X(s,\omega)$.

(ii) Als $\{\omega: T(\omega)<t\} \in F_t$ voor elke $t \in [0,\infty)$ dan geldt

$$\{\omega: T(\omega)\leq t\} = \bigcap_{h>0} \{\omega: T(\omega)<t+h\} \in \bigcap_{h>0} F_{t+h} = F_t$$

vanwege de rechtskontinuïteit van de collectie $\{F_t\}$.

STELLING 2.2.12. *Zij* $\{M(t)\}$ *een kwadratisch integreerbare martingaal horend bij* $\{F_t\}$ *en laten* S *en* T *twee stoptijden zijn t.o.v. de familie* σ*-algebra's* $\{F_t\}$ *zodat* $S \leq T$.
Dan geldt $E(M(T)|F_S) = M(S)$.

BEWIJS. Zie MEYER [19] p.133.

TOEPASSING 2.2.13.

Voor de telprocessen definiëren we:

$T_0 \equiv 0$ voor $n = 1,2,...$

$$T_{n+1} = \begin{cases} \inf\{t: t>T_n(\omega) \text{ en } \vec{N}(t) \neq \vec{N}(T_n)\} & \text{als } \{t: t>T_n(\omega) \text{ en } \vec{N}(t) \neq \vec{N}(T_n)\} \neq \emptyset. \\ \infty & \text{anders} \end{cases}$$

De variabelen $T_1, T_2, ...$ heten de *sprongtijden* van het telproces. Het zijn stoptijden want $T_n(\omega)$ kunnen we schrijven als

$$\begin{cases} \inf\{t: \bar{N}(t) = \sum_{i=1}^{k} N_i(t)>n-1\} & \text{als } \{t: \bar{N}(t)>n-1\} \neq \emptyset. \\ \infty & \text{anders} \end{cases}$$

VOORBEELD VAN EEN TELPROCES 2.1.14.

Zij $\Omega = [0,\infty)$, F = σ-algebra van Borelverzamelingen en P een kansmaat met verdelingsfunctie F en dichtheid f t.o.v. de Lebesgue maat.
Gegeven is een (één-dimensionaal) telproces dat één sprong maakt en wel op tijdstip $X(\omega) = \omega$.

Zij $\mathcal{F}_t$ de σ-algebra voortgebracht door de Borelverzamelingen op $[0,t]$ (genoteerd door $\mathcal{B}([0,t])$), (t,∞) en alle Lebesgue nulverzamelingen van Ω. (Dan is $\mathcal{F}_t$ uitgebreid volledig, rechtskontinu en gelijk aan $\mathcal{N}_t$.)
Het is duidelijk dat X verdelingsfunctie F heeft met dichtheid f en $F(0) = 0$.

Om een indruk te krijgen van het intensiteitsproces berekenen we

$$\lim_{h\downarrow 0} \frac{1}{h} P(N_i(t+h)-N_i(t)=1 \mid \mathcal{F}_t) = \lim_{h\downarrow 0} \frac{1}{h} \frac{F(t+h)-F(t)}{1-F(t)} I_{\{X>t\}}$$

$$= \frac{f(t+)}{1-F(t)} I_{\{X>t\}} \qquad \text{voor P-bijna alle t.}$$

<u>BEWERING</u>. $M(t) = N(t) - \int_0^t \frac{f(s)}{1-F(s)} I_{\{X\geq s\}}\, ds$ *is een martingaal t.o.v.* $\{\mathcal{F}_t\}$.

<u>STAVING</u>.

$$\int_0^t \frac{f(s)}{1-F(s)} I_{\{X\geq s\}} ds = \int_0^{t\wedge X} \frac{f(s)}{1-F(s)} ds = -\log(1-F(t\wedge X)).$$

Zij $t \in [0,\infty)$ en $s \leq t$.
Stel $A \in \mathcal{B}([0,s])$. Als $\omega \in A$ dan $\omega = X(\omega) \leq s < t$.
Dus $N(t) = N(s) = 1$ voor alle $\omega \in A$. Als $\omega \in A$ dan $-\log(1-F(t\wedge X)) =$
$= -\log(1-F(X))$, dus $M(t) = M(s)$ $\forall s \leq t$. Dus voor alle $A \in \mathcal{B}([0,s])$ geldt

$$\int_A M(t)\,dP = \int_A M(s)\,dP \qquad \forall s \leq t.$$

Stel $A = (s,\infty)$.

$$\int_{\{\omega\in(s,\infty)\}} M(s,\omega)\,dP(\omega) = \int_{\{\omega\in(s,\infty)\}} \Big[I_{[X,\infty)}(s)\ (1+\log(1-F(X))) +$$

$$+ I_{[0,X)}(s)\ \log(1-F(s)) \Big]\, dP(\omega)$$

$$= \int_s^\infty \log(1-F(s))\ f(X)\,dX = (1-F(s))\ \log(1-F(s)).$$

Voorts

$$\int_{\{\omega\in(s,\infty)\}} M(t,\omega)\ dP(\omega) = \int_s^\infty I_{[X,\infty)}(t)(1+\log(1-F(X)))f(X)dX$$

$$+ \int_s^\infty I_{[0,X)}(t)\ \log(1-F(t))\ f(X)dX$$

$$= \int_s^t (1+\log(1-F(X)))f(X)dX + \int_t^\infty \log(1-F(t))f(X)dX$$

$$= [-(1-F(X))(\log(1-F(X))]_s^t + (1-F(t))\ \log(1-F(t))$$

$$= (1-F(s))\ \log(1-F(s)).$$

Dus $\quad \int_A M(t,\omega)dP(\omega) = \int_A M(s,\omega)dP(\omega) \quad$ als $A = (s,\infty)$.

Hieruit leidt men eenvoudig af dat deze eigenschap geldt voor alle $A \in F_s = N_s$. Dus $\{M(t)\}$ is een martingaal horend bij $\{F_t\}$.

STELLING 2.2.15. *Stel* $\{F_t\}$ *en* $\{G_t\}$ *zijn voor alle* $t \in [0,\infty)$ *rechtskontinue stijgende rijen σ-algebra's die uitgebreid volledig zijn en* $N_t \subset F_t \subset G_t$ *voor alle* t. *Stel dat het telproces* $\{\vec{N}(t)\}$ *intensiteitsproces* $\{\vec{\Lambda}(t)\}$ *heeft t.o.v.* $\{G_t\}$. *Dan heeft* $\{\vec{N}(t)\}$ *t.o.v.* $\{F_t\}$ *een intensiteitsproces en wel* $E(\vec{\Lambda}(t)|F_{t-})$ *met* $F_{t-} = \bigcup_{s<t} F_s$.

Als $F_t = N_t$ *dan is het intensiteitsproces t.o.v.* $\{N_t\}$ *gelijk aan* $E(\vec{\Lambda}(t)|N_t)$.

BEWIJS. Zie DOLIVO [12] Stelling 3.1.4. p.134-135.

DEFINITIE 2.2.16. Als $\{N(t)\}$ een één-dimensionaal telproces is en $T_1 < T_2 < T_3 < \ldots$ zijn de opeenvolgende sprongtijden met $T_0 \equiv 0$, dan heet $\{N(t)\}$ een *Poisson proces met intensiteit* λ als voor iedere n, $T_1-T_0, T_2-T_1, \ldots, T_n-T_{n-1}$ onafhankelijk en identiek verdeeld zijn, elk met dichtheid $f(t) = \lambda e^{-\lambda t}$.

De volgende stelling geeft het verband tussen stochastische integralen en Lebesgue-Stieltjes integralen.

Stel dat $\{H(t)\}$ een voorspelbaar proces is en dat $\{N(t)\}$ een één-dimensionaal telproces is met intensiteitsproces $\{\Lambda(t)\}$ t.o.v. $\{F_t\}$ zodanig dat $M(t) = N(t) - \int_0^t \Lambda(s)ds$ een kwadratisch integreerbare martingaal is. Nu is $\int_0^t H(s)dM(s)$ als stochastische integraal gedefinieerd indien $E(\int_0^\infty H^2(s)d\langle M,M\rangle(s)) < \infty$. Als aan deze eis voldaan is, geldt in het algemeen niet dat deze stochastische integraal gelijk is aan het verschil van twee Lebesgue-Stieltjes integralen: $\int_0^t H(s)dN(s) - \int_0^t H(s)\Lambda(s)ds$. Onder de voorwaarden van de volgende stelling is dit wel het geval.

STELLING 2.2.17. *Als*

$$E\left(\int_0^\infty |H(s)|\ dN(s)\right) < \infty \ \textit{én}\ E\left(\int_0^\infty |H(s)|\ \Lambda(s)ds\right) < \infty$$

dan geldt dat

$$\int_0^t H(s)dM(s) = \int_0^t H(s)dN(s) - \int_0^t H(s)\ \Lambda(s)ds$$

met kans 1 *voor iedere* $t \in [0,\infty)$.

BEWIJS. Zie DOLÉANS-DADE & MEYER [11] bewering 3 p.90-91.

2.3. ZWAKKE KONVERGENTIE VAN STOCHASTISCHE INTEGRALEN

De stelling die we in deze paragraaf gaan behandelen is geldig op het interval $[0,\tau]$, waarbij τ de waarde ∞ mag aannemen. Het bewijs van stelling 2.3.1. is te vinden in AALEN [3].

Stel $\{\vec{N}_n(t)\} = \{(N_{1n}(t),\ldots,N_{kn}(t))\}$ is een k-dimensionaal telproces op $[0,\tau]$. Voor $n = 1,2,\ldots$ is dit telproces gedefinieerd op $(\Omega_n, \mathcal{A}_n, P_n)$.
Stel dat t.o.v. σ-algebra's $\mathcal{F}_{tn} \subset \mathcal{A}_n$, met $t \in [0,\tau]$, $\{\vec{N}_n(t)\}$ intensiteitsproces $\{\vec{\Lambda}_n(t)\}$ heeft en dat $E(N_{in}(\tau)) < \infty$ voor alle i en n. Definieer $M_{in}(t) = N_{in}(t) - \int_0^t \Lambda_{in}(s)ds$ voor $i = 1,2,\ldots,k$ (dit zijn kwadratisch integreerbare martingalen t.o.v. $\{\mathcal{F}_{tn}\}$).
Stel dat $\{\vec{H}_n(t)\}$ k-dimensionale voorspelbare processen op $[0,\tau]$ zijn zodanig dat

$$E\left\{\int_0^\tau H_{in}^2(s)\ \Lambda_{in}(s)ds\right\} < \infty$$

$$E\left\{\int_0^\tau |H_{in}(s)|\ dN_{in}(s)\right\} < \infty \qquad \begin{array}{l} i = 1,2,\ldots,k \\ \\ n = 1,2,\ldots \end{array}$$

$$E\left\{\int_0^\tau |H_{in}(s)|\ \Lambda_{in}(s)ds\right\} < \infty$$

Onder deze voorwaarden zijn de komponent processen van $\{\vec{Y}_n(t)\}$ voor iedere $t \in [0,\tau]$ gedefinieerd door $Y_{in}(t) = \int_0^t H_{in}(s)dM_{in}(s)$, voor $i = 1,2,\ldots,k$, Lebesgue-Stieltjes integralen (zie Stelling 2.2.17.) en tevens stochastische integralen. Bovendien zijn het kwadratisch integreerbare martingalen met

$$\langle Y_{in}, Y_{in}\rangle(t) = \int_0^t H_{in}^2(s)\Lambda_{in}(s)ds \qquad i = 1,2,\ldots,k$$

$$\langle Y_{in}, Y_{jn}\rangle(t) = 0 \qquad i \neq j.$$

(zie Stelling 2.1.29. en gevolg 2.1.33.).

Stel dat $D[0,\tau]$ de ruimte van functies op $[0,\tau]$ is die rechtskontinu zijn en linkerlimieten hebben en verder dat $D^k[0,\tau]$ het cartesisch produkt is van k kopieën van $D[0,\tau]$.

Op $D^k[0,\tau]$ noteren we zwakke konvergentie mbt. de produkt Skorohod topologie (zie BILLINGSLEY [4] p.111 e.v.) door $\xrightarrow{zw}$ (k=1,2,...).

We beschouwen $\{\vec{Y}_n(t)\}$ als een stochastisch element van $D^k[0,\tau]$.

Definieer

$$\overline{N}_n(t) = \sum_{i=1}^{k} N_{in}(t)$$

$$\overline{\Lambda}_n(t) = \sum_{i=1}^{k} \Lambda_{in}(t)$$

en $\{W_1(t)\},\ldots,\{W_k(t)\}$ als k onafhankelijke Wiener processen, $t \in [0,\tau]$, op een kansruimte (Ω,A,P).

STELLING 2.3.1. *Laat aan de volgende voorwaarden voldaan zijn:*

(a) Er bestaan niet-negatieve Borel-meetbare kwadratisch integreerbare functies $g_i\,(i=1,2,\ldots,k)$ *op* $[0,\tau]$ *zodanig dat*

$$\int_0^t H_{in}^2(s)\ \Lambda_{in}(s)ds \xrightarrow{P} \int_0^t g_i^2(s)ds \qquad \textit{voor } i = 1,2,\ldots,k \qquad \forall t \in (0,\tau]$$

($\xrightarrow{P}$ *betekent konvergentie in waarschijnlijkheid).*

$$(b)\qquad E\left\{\sum_{i=1}^{k}\int_0^\tau H_{in}^2(s)\ I_{\{|H_{in}(s)|>\varepsilon\}}dN_{in}(s)\right\} \to 0 \qquad \textit{als } n \to \infty \qquad \forall \varepsilon > 0.$$

(c) Voor iedere n *bestaat er een niet-dalend proces* $\{Z_n(t)\}$ *en een stochastische grootheid* X_n *gedefinieerd op* (Ω_n,A_n,P_n) *zodanig dat*

$$E(Z_n(\tau)-Z_n(0)) < \infty$$

$$E(X_n\overline{N}_n(\tau)) < \infty$$

$$E(X_n\int_0^\tau \overline{\Lambda}_n(s)ds) < \infty$$

en (i) $|H_{in}^2(t)\Lambda_{in}(t) - H_{in}^2(s)\ \Lambda_{in}(s)| \le Z_n(t) - Z_n(s)$ *voor alle* $0 \le s \le t \le \tau$ *en* $\forall i,n$.

(ii) $H_{in}^2(t)\ \Lambda_{in}(t) \le X_n \qquad \forall i,n,t.$

Dan geldt $\{\vec{Y}_n(t)\} \xrightarrow{zw} \{\vec{Y}(t)\}$ *waarin* $\vec{Y}(t) = (Y_1(t),\ldots,Y_k(t))$ *en* $Y_i(t) = \int_0^t g_i(s)dW_i(s)$ *voor* $i = 1,\ldots,k.$

OPMERKINGEN 2.3.2.

A. Als $$E\left\{ \int_0^\tau H_{in}^4(s)\, I_{\{|H_{in}(s)|>\varepsilon\}}\, \Lambda_{in}(s)ds \right\} < \infty \qquad \forall i,n$$

en $$E\left\{ \int_0^\tau H_{in}^2(s)\, I_{\{|H_{in}(s)|>\varepsilon\}}\, dN(s) \right\} < \infty \qquad \forall i,n$$

dan is voorwaarde (b) equivalent met

$$E\left\{ \sum_{i=1}^{k} \int_0^\tau H_{in}^2(s)\, I_{\{|H_{in}(s)|>\varepsilon\}}(s)\, \Lambda_{in}(s)ds \right\} \to 0$$

als $n \to \infty \quad \forall \varepsilon > 0$.

B. $\{\vec{Y}(t)\}$ is gedefinieerd omdat $\int_0^\tau g_i^2(s)ds < \infty$ voor $i = 1,2,\ldots,k$ en omdat $g_i(t)$ voor $i = 1,2,\ldots,k$ Borel-meetbaar is. (Een Borel-meetbare functie is voorspelbaar (zie DELLACHERIE [8] stelling 4 p. 68).)

C. Voorwaarde (a) kunnen we als volgt verklaren.

$$E(Y_{in}^2(t)) = E(\langle Y_{in}, Y_{in}\rangle(t)) = E\{ \int_0^t H_{in}^2(s)\Lambda_{in}(s)ds\}$$

$$E(Y_i^2(t)) = E(\langle Y_i, Y_i\rangle(t)) = E\{ \int_0^t g_i^2(s)ds\}.$$

Voorwaarde (a) is nodig om in de limiet een deterministisch voorspelbaar groeiend proces te krijgen.

D. Voorwaarde (b) is een soort Lindeberg voorwaarde.
De voorwaarde impliceert dat in de limiet $H_{in}(t)$ klein wordt dus de sprongen van $Y_{in}(t)$ moeten klein worden.
Gekombineerd met voorwaarde (a) moet dus gelden voor $g(t) \neq 0$ dat $\Lambda_{in}(t)$ (en dus ook $N_{in}(t)$) voor $i = 1,2,\ldots,k$ groot moet worden.

E. Voorwaarde (c) is lastig. In het volgende hoofdstuk hebben we voor $i = 1,2,\ldots,k$ dat $\Lambda_{in}(t) = \lambda_i(t)K_{in}(t)$ met $\lambda_i(t)$ een functie in t en $\{K_{in}(t)\}$ een speciaal stochastisch proces. In die gevallen moeten we veronderstellen dat $\lambda_i(t)$ van begrensde variatie is op eindige intervallen. Deze eis is vervuld als $\lambda_i(t)$ één keer kontinu differentieerbaar is.

In onze toepassingen kunnen wij volstaan met een verzwakking van voorwaarde (c). Deze verzwakking luidt: $\lambda_i(t)$ is begrensd op $[0,\tau]$ en $\int_0^\tau \lambda_i(t)dt < \infty$ voor $i = 1,2,\ldots,k$.

Tot slot van dit hoofdstuk nog een eigenschap van zwakke konvergentie.

Stel A is een interval bevat in $\mathbb{R}_+$. Definieer $D(A)$ als de ruimte van functies x op A die rechtskontinu zijn en linkerlimieten hebben.
Stel dat voor $n = 1,2,\ldots$ $\{Z_n(t)\}$ en $\{Z(t)\}$ stochastische processen zijn voor $t \in A$.
Stel dat $\{Z_n(t,\omega)\}$ als functie van t een element is van $D(A)$, $n = 1,2,\ldots$ en laat $Z(t)$ met kans 1 kontinue paden hebben.

<u>STELLING 2.3.3.</u> $\{Z_n(t)\}_{t\in A} \overset{zw}{\rightarrow} \{Z(t)\}_{t\in A}$ *dan en slechts dan als* $\{Z_n(t)\}_{t\in K} \overset{zw}{\rightarrow} \{Z(t)\}_{t\in K}$ *voor ieder kompakt interval* $K \subset A$.

<u>BEWIJS</u>. Zie VERVAAT [27] Stelling 1.3.10. p.18-20.

In het eerste hoofdstuk hebben we drie toetsen behandeld. In het volgende hoofdstuk zullen we een aantal eigenschappen van deze toetsen afleiden met behulp van de theorie van telprocessen.

<u>Appendix</u>

<u>BEWIJS</u> van de twee eigenschappen.
Voor $i = 1,2,\ldots,k$ geldt

$$
\begin{aligned}
U_i(t) &= \lim_{h\downarrow 0} \frac{1}{h} E(N_i(t+h) - N_i(t) \mid F_t) \\
&= \lim_{h\downarrow 0} \frac{1}{h} E\left[(N_i(t+h) - N_i(t))\, I_{\{N_i(t+h)-N_i(t)>1\}} \,\middle|\, F_t\right] + \\
&\quad \lim_{h\downarrow 0} \frac{1}{h} E\left[(N_i(t+h) - N_i(t))\, I_{\{N_i(t+h)-N_i(t)\le 1\}} \,\middle|\, F_t\right] \\
&= \lim_{h\downarrow 0} \frac{1}{h} E\left[(N_i(t+h) - N_i(t))\, I_{\{N_i(t+h)-N_i(t)>1\}} \,\middle|\, F_t\right] + U_i(t).
\end{aligned}
$$

Dus
$$\lim_{h\downarrow 0} \frac{1}{h} E\left[(N_i(t+h) - N_i(t))\, I_{\{N_i(t+h)-N_i(t)>1\}} \,\middle|\, F_t\right] = 0.$$

$$\lim_{h\downarrow 0} \frac{1}{h} P\left[N_i(t+h)-N_i(t)>1 \mid F_t\right] = \lim_{h\downarrow 0} \frac{1}{h} \left[1-P(N_i(t+h)-N_i(t)\le 1 \mid F_t)\right]$$

$$= \lim_{h\downarrow 0} \frac{1}{h}\left[1-P(N_i(t+h)-N_i(t)=0 \mid F_t) - P(N_i(t+h)-N_i(t)=1 \mid F_t)\right]$$

$$= U_i(t) - U_i(t) = 0. \qquad \square$$

<u>BEWIJS VAN STELLING 2.2.8.</u>

We gaan eerst eis (a) na. Voor $i = 1,2,\dots,k$ geldt

$$\lim_{h\downarrow 0} E\left[\frac{1}{h}(N_i(t+h)-N_i(t)) \mid F_t\right] = \lim_{h\downarrow 0} E\left[\frac{1}{h}\int_t^{t+h} \Lambda_i^*(s)ds \mid F_t\right]$$

$$= E\left[\lim_{h\downarrow 0} \frac{1}{h}\int_t^{t+h} \Lambda_i^*(s)ds \mid F_t\right] = E(\Lambda_i^*(t) \mid F_t) = \Lambda_i^*(t)$$

(wegens voorwaardelijke gedomineerde konvergentie stelling, zie LOÈVE [17] p.348). Om (b) te bewijzen nemen we i en t vast.

Stel $S > t$ is het tijdstip van de eerste sprong na tijdstip t van $N_i(s)$. S is een stoptijd.

Het proces $I_{(t,S]}$ hoort bij $\{F_t\}$ en is linkskontinu (dus voorspelbaar).

De stochastische integraal is gedefinieerd als $\int_t^{t+h} I_{(t,S]}(u)dM_i(u)$ met $M_i(t) = N_i(t) - \int_0^t \Lambda_i(s)ds$.

Dit proces met h als tijdsvariabele is een martingaal, dus geldt dat

$$\lim_{h\downarrow 0} \frac{1}{h} P(N_i(t+h)-N_i(t)=1 \mid F_t) = \lim_{h\downarrow 0} \frac{1}{h} E\left[\int_t^{t+h} I_{(t,S]}(u)dN_i(u) \mid F_t\right]$$

$$= \lim_{h\downarrow 0} \frac{1}{h} E\left[\int_t^{t+h} I_{(t,S]}(u)\Lambda_i^*(u)du \mid F_t\right]$$

$$= E\left(\lim_{h\downarrow 0} \frac{1}{h}\int_t^{t+h} I_{(t,S]}(u)\Lambda_i^*(u)du \mid F_t\right) = E(\Lambda_i^*(t) \mid F_t) = \Lambda_i^*(t)$$

(wegens voorwaardelijke gedomineerde konvergentie stelling). Verder geldt dat

$$0 \le \frac{1}{h}\left[1-P(N_i(t+h)-N_i(t)=0 \mid F_t)\right] - \frac{1}{h} P(N_i(t+h)-N_i(t)=1 \mid F_t)$$

$$\le \frac{1}{h} E(N_i(t+h)-N_i(t) \mid F_t) - \frac{1}{h} P(N_i(t+h)-N_i(t)=1 \mid F_t)$$

$$\longrightarrow \Lambda_i^*(t) - \Lambda_i^*(t) = 0 \qquad \text{als } h\downarrow 0. \qquad \square$$

HOOFDSTUK 3

3.1. HET MODEL

Allereerst beschouwen we een steekproef van onderling onafhankelijk en identiek verdeelde stochastische grootheden.
Stel $X_1, X_2, \ldots, X_n$ zijn o.o. identiek verdeelde stochastische grootheden op een kansruimte (Ω_1, A_1, P_1). De σ-algebra A_1 is volledig met betrekking tot de kansmaat P_1.
Stel X_1 heeft verdelingsfunctie F waarvoor geldt dat $F(0) = 0$, $F(\tau) = 1$ en $F(t) < 1$ voor alle $t < \tau$, waarbij $\tau \in [0,\infty]$. Verder heeft F een dichtheid f t.o.v. de Lebesgue maat.
Definieer:

$$N_i(t) = N_i(t,\omega_1) = I_{\{X_i \leq t\}}(\omega_1) \qquad \text{voor } i = 1,2,\ldots,n$$

$$\vec{N}(t) = (N_1(t), N_2(t), \ldots, N_n(t)).$$

Stel dat $\{N_{1,t}\}$ de uitgebreid volledige deel σ-algebra's van A_1 zijn, voortgebracht door $N_i(s)$ voor alle $s \leq t$ en $i = 1,2,\ldots,n$, waarbij $t \in [0,\tau]$.
$\{\vec{N}(t)\}$ is een telproces met intensiteitsproces $\{\vec{\Lambda}(t)\}$ t.o.v. $\{N_{1,t}\}$ gegeven door

$$\Lambda_i(t) = \lambda(t).I_{\{X_i \geq t\}} \qquad \text{en} \qquad \lambda(t) = \begin{cases} \dfrac{f(t)}{1-F(t)} & 0 \leq t < \tau \\ 0 & t = \tau \end{cases}$$

$$\text{voor } i = 1,2,\ldots,n$$

(zie voorbeeld 2.2.14.).

Definieer $\bar{N}(t) = \sum_{i=1}^n N_i(t)$ en $\bar{\Lambda}(t) = \sum_{i=1}^n \Lambda_i(t) = \lambda(t)(n-\bar{N}(t-))$ dan is

$\{\bar{N}(t)\}$ een telproces met intensiteitsproces $\{\bar{\Lambda}(t)\}$ t.o.v. $\{N_{1,t}\}$.

Dit model gaan we nu uitbreiden tot *een model voor stochastische censurering*.

Stel $X_1, X_2, \ldots, X_n$ zijn als voorheen.

Stel (Ω_2, A_2, P_2) is een andere kansruimte en (Ω, A, P) de volledige produktruimte.

F_t is de produkt σ-algebra voortgebracht door $N_{1,t} \times A_2$ en aangevuld met alle P-nulverzamelingen uit A $(t \in [0,\tau])$.

F_t is rechtskontinu en $\vec{N}(t)$ en $\vec{\Lambda}(t)$ worden op Ω gedefinieerd door $\vec{N}(t,\omega) = \vec{N}(t,\omega_1)$ en $\vec{\Lambda}(t,\omega) = \vec{\Lambda}(t,\omega_1)$ met $\omega = (\omega_1, \omega_2)$.

Uit het voorgaande weten we dat

$$M_i(t,\omega_1) = N_i(t,\omega_1) - \int_0^t \Lambda_i(s,\omega_1)ds \qquad \text{voor } i = 1,2,\ldots,n$$

orthogonale kwadratisch integreerbare martingalen t.o.v. $\{N_{1,t}\}$ zijn.

<u>BEWERING 3.1.1.</u> $M_i(t,\omega)$ *zijn voor* $i = 1,2,\ldots,n$ *orthogonale kwadratisch integreerbare martingalen t.o.v.* $\{F_t\}$ *als* $M_i(t,\omega) \overset{\text{def}}{=} M_i(t,\omega_1)$. *Dus* $\{\vec{N}(t)\}$ *is een telproces met intensiteitsproces* $\{\vec{\Lambda}(t)\}$ *t.o.v.* $\{F_t\}$.

<u>BEWIJS</u> (schets). Neem $A \in N_{1,s}$ en $B \in A_2$. Dan geldt m.b.v. de stelling van Fubini

$$\int_{A\times B} M_i(t,\omega)dP(\omega) = \int_A \int_B M_i(t,\omega_1)dP_2(\omega_2)dP_1(\omega_1) =$$

$$= \int_A P_2(B)M_i(t,\omega_1)dP_1(\omega_1) = \int_A P_2(B)M_i(s,\omega_1)dP_1(\omega_1) =$$

$$= \int_A \int_B M_i(s,\omega_1)dP_2(\omega_2)dP_1(\omega_1) = \int_{A\times B} M_i(s,\omega)dP(\omega)$$

$$\forall s \le t.$$

Hiermee hebben we de bewering aangetoond voor eventualiteiten die de σ-algebra $\{F_t\}$ voortbrengen. Het is uit te breiden voor de algebra die deze eventualiteiten voortbrengen en dus voor de σ-algebra $\{F_t\}$. □

Het proces $\{\vec{J}(t)\} = \{(J_1(t), J_2(t), \ldots, J_n(t))\}$ is een proces dat hoort bij $\{F_t\}$, linkskontinu is en rechterlimieten heeft. De komponent processen nemen alleen de waarden 0 en 1 aan.
In deze syllabus bespreken we alleen rechter-censurering. In dat geval eisen we dat als $J_i(t) = 0$ dan $J_i(s) = 0$ voor alle $s \geq t$. In het vervolg zal deze eis gelden. Als een komponent proces van $\vec{J}(t)$ gelijk aan 0 wordt, betekent dit dat vanaf dat moment de waarneming van X_i gecensureerd wordt.

Het *censureringsproces* $\{\vec{N}^*(t)\}$ wordt nu gedefinieerd door

$$N_i^*(t) = \int_0^t J_i(s)dN_i(s) \qquad \text{voor } i = 1,2,\ldots,n.$$

$N_i^*(t) = 1$ geldt dan en slechts dan als $X_i \leq t$ en $J_i(X_i) = 1$.
Nu is

$$M_i(t) = N_i(t) - \int_0^t \Lambda_i(s)ds$$

een $\{F_t\}$-martingaal en $\{\vec{J}(t)\}$ een voorspelbaar proces dus

$$M_i^*(t) = \int_0^t J_i(s)dM_i(s) \qquad \text{voor } i = 1,2,\ldots,n$$

zijn $\{F_t\}$-martingalen die orthogonaal en kwadratisch integreerbaar zijn.
Volgens stelling 2.2.17. geldt dat de stochastische integraal gelijk is aan het verschil van twee Lebesgue-Stieltjes integralen nl.

$$M_i^*(t) = \int_0^t J_i(s)dM_i(s) = \int_0^t J_i(s)dN_i(s) - \int_0^t J_i(s)\Lambda_i(s)ds$$

$$\text{voor } i = 1,2,\ldots,n.$$

Dus $\{\vec{N}^*(t)\}$ is een telproces met intensiteitsproces $\{\vec{\Lambda}^*(t)\}$ gegeven door

$$\Lambda_i^*(t) = \Lambda_i(t)J_i(t) \qquad \text{voor } i = 1,2,\ldots,n$$

t.o.v. $\{F_t\}$.
Volgens voorgaande geldt voor $i = 1,2,\ldots,n$ dat

$$\Lambda_i^*(t,\omega) = \lambda(t)I_{\{X_i \geq t\}}(\omega)J_i(t,\omega).$$

Definieer $\bar{N}^*(t) = \sum_{i=1}^n N_i^*(t)$ en $B(t) = \sum_{i=1}^n I_{\{X_i \geq t\}} J_i(t)$.
Dan is $\{\bar{N}^*(t)\}$ een telproces met intensiteitsproces $\{\lambda(t)B(t)\}$ t.o.v. $\{\mathcal{F}_t\}$.
Stel dat $U_1, U_2, \ldots, U_n$ gedefinieerd op Ω, censurerende variabelen zijn zodanig dat $I_{\{U_i \geq t\}}$ $\{\mathcal{F}_t\}$-meetbaar zijn.
Definieer $J_i(t) = I_{\{U_i \geq t\}}$. Dan is $\vec{J}(t)$ linkskontinu met rechterlimieten en $\{\mathcal{F}_t\}$-meetbaar.
Definieer $\tilde{X}_i = X_i \wedge U_i$ voor $i = 1,\ldots,n$. Dan is $N_i^*(t) = I_{\{\tilde{X}_i \leq t \text{ en } \tilde{X}_i = X_i\}}$ voor $i = 1,\ldots,n$ en $\Lambda_i^*(t) = \lambda(t) I_{\{\tilde{X}_i \geq t\}}$.
De eisen die we aan de ruimte Ω hebben opgelegd en de eis dat $\vec{J}(t)$ hoort bij $\{\mathcal{F}_t\}$ leveren beperkingen op van de censureringstypen (dwz. een model waarin $X_1, X_2, \ldots, X_n$ o.o. identiek verdeelde stochastische grootheden zijn zodanig dat men alleen de stochastische grootheden $X_i \wedge U_i$ en de indicatorvariabelen $I_{\{X_i \leq U_i\}}$ waarneemt voor zekere stochastische grootheden U_i voor $i = 1,2,\ldots,n$) die wij hiermee kunnen behandelen.

<u>VOORBEELDEN 3.1.2.</u>

(i) $U_i \equiv u_i$ dwz. alle objekten worden ieder op een vast tijdstip gecensureerd (gevallen (i) en (ii) uit de inleiding van hoofdstuk 1).

(ii) $\vec{U}$ onafhankelijk van $\vec{X}$ maar met een willekeurige simultane verdeling van de komponenten van $\vec{U}(t)$. (In dit geval definiëren we $\vec{U}$ op Ω_2 en vervolgens op Ω door $U_i(\omega) = U_i(\omega_2)$ voor $i = 1,2,\ldots,n$. Dan is $\vec{J}(t)$ meetbaar t.o.v. $\mathcal{F}_t = \mathcal{N}_{1,t} \times \mathcal{A}_2$. Gevallen (iii) en (iv) uit de inleiding van hoofdstuk 1.)

(iii) $\vec{J}(t)$ is een functie van $\{\vec{N}(s);\ s \in [0,t]\}$ voor alle t. Bijvoorbeeld $U_i \equiv X_{(r)}$ voor alle i.

3.2. DE TOETSEN VAN GEHAN, EFRON EN COX

Stel $X_1, X_2, \ldots, X_m$ zijn o.o. en identiek verdeelde stochastische grootheden met uitvalsfunctie $\lambda_1(x)$. Definieer $\tilde{X}_i = X_i \wedge U_i$ en $\Delta_i = I_{\{\tilde{X}_i = X_i\}}$ voor $i = 1,2,\ldots,m$.
Stel $Y_1, Y_2, \ldots, Y_n$ zijn o.o. en identiek verdeelde stochastische grootheden met uitvalsfunctie $\lambda_2(y)$. Definieer $\tilde{Y}_j = Y_j \wedge V_j$ en $E_j = I_{\{\tilde{Y}_j = Y_j\}}$ voor $j = 1,2,\ldots,n$.
Definieer:
$$N_1(t) = \#\{i:\ \tilde{X}_i \leq t \text{ en } \Delta_i = 1\}$$
$$N_2(t) = \#\{j:\ \tilde{Y}_j \leq t \text{ en } E_j = 1\}$$
$$B_1(t) = \#\{i:\ \tilde{X}_i \geq t\}$$
$$B_2(t) = \#\{j:\ \tilde{Y}_j \geq t\}.$$

Stel dat er σ-algebra's $\{\mathcal{F}_t\}$ bestaan zodanig dat $\{\vec{N}(t)\} = \{(N_1(t), N_2(t))\}$ een telproces is met intensiteitsproces $\{(\lambda_1(t)B_1(t), \lambda_2(t)B_2(t))\}$ t.o.v. de σ-algebra's $\{\mathcal{F}_t\}$.

De nulhypothese H_0 is $\lambda_1(t) = \lambda_2(t)$ voor alle $t \in [0,\tau)$ met $\tau \in [0,\infty]$.
Definieer $J(t) = I_{\{B_1(t)>0 \text{ én } B_2(t)>0\}}$ voor $t \in [0,\tau)$.
In het vervolg nemen we voor $\tau = \infty$ tenzij anders vermeld.
Definieer

$$\hat{G}_i(t) = \int_0^t \frac{J(s)}{B_i(s)}\, dN_i(s) \qquad \text{voor } i = 1,2$$

$$G_i(t) = \int_0^t J(s)\lambda_i(s)ds \qquad \text{voor } i = 1,2.$$

<u>OPMERKING 3.2.1</u>. Als $B_i(t) = 0$ dan definiëren we $\frac{J(t)}{B_i(s)} = 0$ voor $i = 1,2$.
Nu is $\left\{\frac{J(s)}{B_i(s)}\right\}$ voor $i = 1,2$ een begrensd linkskontinu proces dat hoort bij $\{\mathcal{F}_t\}$.
Dus $\int_0^t \frac{J(s)}{B_i(s)}\, dM_i(s)$ is een kwadratisch integreerbare martingaal voor $i = 1,2$ met $M_i(t) = N_i(t) - \int_0^t \lambda_i(s)ds$.
Volgens stelling 2.2.17. geldt $\int_0^t \frac{J(s)}{B_i(s)}\, dM_i(s) = \hat{G}_i(t) - G_i(t)$.
Dus $\{\hat{G}_1(t) - G_1(t)\}$ is een kwadratisch integreerbare martingaal voor $i = 1,2$.
Hieruit volgt dat $\{\hat{G}_2(t) - G_2(t) - \hat{G}_1(t) + G_1(t)\}$ een kwadratisch integreerbare martingaal is.
Onder H_0 geldt dat $G_1(t) = G_2(t)$ zodat onder H_0 geldt dat $\hat{G}_2(t) - \hat{G}_1(t)$ een kwadratisch integreerbare martingaal is.

Stel $\{K(t)\}$ is een niet-negatief linkskontinu proces horend bij $\{\mathcal{F}_t\}$ en zodanig dat

$$E\left(\int_0^\infty K^2(s)\, \frac{J(s)}{B_i(s)}\, \lambda_i(s)ds\right) < \infty \qquad \text{voor } i = 1,2.$$

Definieer

$$Z(t) = \int_0^t K(s)\, \frac{J(s)}{B_2(s)}\, dN_2(s) - \int_0^t K(s)\, \frac{J(s)}{B_1(s)}\, dN_1(s) =$$

$$= \int_0^t K(s)\, \frac{J(s)}{B_2(s)}\, dM_2(s) - \int_0^t K(s)\, \frac{J(s)}{B_1(s)}\, dM_1(s) +$$

$$+ \int_0^t (\lambda_2(s) - \lambda_1(s))J(s)ds.$$

Een andere schrijfwijze is

$$Z(t) = \int_0^t K(s)J(s)d(\hat{G}_2(s)-\hat{G}_1(s)).$$

OPMERKING 3.2.2. De voorwaarde op het proces $\{K(t)\}$ impliceert dat $Z(t)$ als stochastische integraal gedefinieerd is. $\{Z(t)\}$ is onder H_0 een kwadratisch integreerbare martingaal. Als bijvoorbeeld $\{K(t)\}$ begrensd is met kans 1 door een konstante is $\{Z(t)\}$ tevens als Lebesgue-Stieltjes integraal gedefinieerd.

$$\begin{aligned} E(Z(\infty)) &= E\left(\int_0^\infty K(s)J(s)d(\hat{G}_2(s)-\hat{G}_1(s))\right) = \\ &= E\left(\int_0^\infty K(s)J(s)d(G_2(s)-G_1(s))\right) = \\ &= E\left(\int_0^\infty K(s)J(s)(\lambda_2(s)-\lambda_1(s))ds\right) = \\ &= \int_0^\infty E(K(s)J(s))(\lambda_2(s)-\lambda_1(s))ds. \end{aligned}$$

We zien dat $Z(\infty)$ een geschikte toetsingsgrootheid lijkt voor de alternatieve hypothese $\lambda_2(s) > \lambda_1(s)$ voor alle $s \geq 0$. Immers onder H_0 geldt $E(Z(\infty)) = 0$ en onder het alternatief $EZ(\infty) > 0$.

Als $\{K(t)\}$ bovendien niet-stijgend en begrensd is dan

$$\begin{aligned} E(Z(\infty)) &= \left[E(K(s)J(s))\log\left(\frac{1-F_1(s)}{1-F_2(s)}\right)\right]_0^\infty + \int_0^\infty \log\left(\frac{1-F_1(s)}{1-F_2(s)}\right)d(-E(K(s)J(s)) = \\ &= \int_0^\infty \log\left(\frac{1-F_1(s)}{1-F_2(s)}\right)d(-E(K(s)J(s)). \end{aligned}$$

Onder deze voorwaarden lijkt $Z(\infty)$ een redelijke toetsingsgrootheid voor de algemenere alternatieve hypothese $F_2(t) > F_1(t)$ voor alle t.

We zullen nu drie verschillende keuzen van het proces $\{K(t)\}$ bespreken. Voor alle $t \in [0,\infty)$ definiëren we:

(i) $K_1(t) = B_1(t)B_2(t)$.

(ii) $K_2(t) = \hat{S}_1(t-)\hat{S}_2(t-)$ waarbij $\hat{S}_1(t-)$ de gemodificeerde produkt-limiet schatter is van $P(X_i \geq t)$ en $\hat{S}_2(t-)$ de gemodificeerde produkt-limiet schatter is van $P(Y_j \geq t)$. De *gemodificeerde produkt-limiet schatter* is *voor alle waarden* van het argument gedefinieerd door de eerste regel van definitie 1.3.1.

$$\text{(iii)}\quad K_3(t) = \begin{cases} \dfrac{B_1(t)B_2(t)}{B_1(t)+B_2(t)} = (B_1^{-1}(t) + B_2^{-1}(t))^{-1} & \text{als } B_1(t) + B_2(t) > 0 \\ 0 & \text{als } B_1(t) + B_2(t) = 0. \end{cases}$$

OPMERKING 3.2.3. $\{K_i(t)J(t)\}$ is niet-negatief, linkskontinu, niet-stijgend en begrensd voor $i = 1,2,3$.

Definieer voor $i = 1,2$ en 3

$$Z_i(t) = \int_0^t K_i(s)J(s)\,d(\hat{G}_2(s)-\hat{G}_1(s)).$$

STELLING 3.2.4. $Z_1(\infty)$ *is de toetsingsgrootheid van Gehan.*

BEWIJS.

$$\begin{aligned} Z_1(\infty) &= \int_0^\infty K_1(s)J(s)\,d(\hat{G}_2(s)-\hat{G}_1(s)) = \\ &= \int_0^\infty B_1(s)B_2(s)J(s)\left(\frac{J(s)}{B_2(s)}\,dN_2(s) - \frac{J(s)}{B_1(s)}\,dN_1(s)\right) = \\ &= \int_0^\infty B_1(s)J(s)\,dN_2(s) - \int_0^\infty B_2(s)J(s)\,dN_1(s) = \\ &= \int_0^\infty B_1(s)\,dN_2(s) - \int_0^\infty B_2(s)\,dN_1(s) = \\ &= W \text{ (zie definitie 1.2.1.).} \qquad \square \end{aligned}$$

STELLING 3.2.5. $Z_3(\infty)$ *is de toetsingsgrootheid van Cox.*

BEWIJS.

$$\begin{aligned} Z_3(\infty) &= \int_0^\infty \frac{B_1(s)B_2(s)}{B_1(s)+B_2(s)}\,J(s)\,d(\hat{G}_2(s)-\hat{G}_1(s)) = \\ &= \int_0^\infty J(s)\,\frac{B_1(s)}{B_1(s)+B_2(s)}\,dN_2(s) - \int_0^\infty J(s)\,\frac{B_2(s)}{B_1(s)+B_2(s)}\,dN_1(s) = \end{aligned}$$

$$= \int_0^\infty \frac{B_1(s) + B_2(s) - B_2(s)dN_2(s)}{B_1(s) + B_2(s)} - \int_0^\infty \frac{B_2(s)}{B_1(s) + B_2(s)}\, dN_1(s) =$$

$$= \int_0^\infty dN_2(s) - \int_0^\infty \frac{B_2(s)}{B_1(s) + B_2(s)}\, d(N_1(s) + N_2(s)) =$$

$$= N_2(\infty) - \int_0^\infty \frac{B_2(s)}{B_1(s) + B_2(s)}\, d(N_1(s) + N_2(s)) =$$

$$= N_1 - \sum_{i=1}^{s} \Pi_i = U(0) \text{ (zie definitie 1.5.4.).} \qquad \Box$$

<u>OPMERKING 3.2.6</u>. Hieronder gaan we na het verband tussen de toetsingsgrootheid van Efron met $\frac{1}{2} + \frac{1}{2}Z_2(\infty)$.

Definieer $\Delta X(t) = X(t+) - X(t-)$.

Dan

$$\Delta N_i(t) = 1 \iff \hat{S}_i(t) = \hat{S}_i(t-)\left(1 - \frac{1}{B_i(t)}\right) \iff \Delta\hat{S}_i(t) = -\frac{\hat{S}_i(t-)}{B_i(t)}$$

voor $i = 1,2$.

Hieruit volgt

$$Z_2(\infty) = \int_0^\infty J(s)\hat{S}_1(s-)\hat{S}_2(s-)\left(\frac{dN_2(s)}{B_2(s)} - \frac{dN_1(s)}{B_1(s)}\right) =$$

$$= \int_0^\infty J(s)(-\hat{S}_1(s-))d\hat{S}_2(s) - \int_0^\infty J(s)(-\hat{S}_2(s-))d\hat{S}_1(s) =$$

$$= \int_0^\infty J(s)\hat{S}_2(s-)d\hat{S}_1(s) - \int_0^\infty J(s)\hat{S}_1(s-)d\hat{S}_2(s).$$

De toetsingsgrootheid van Efron was in definitie 1.4.1. gedefinieerd als

$$W = -\int_0^\infty S_m(s)dT_n(s) \quad \text{met} \quad S_m(s) = \begin{cases} \hat{S}_1(s) & \text{als } B_1(s) > 0 \\ 0 & \text{als } B_1(s) = 0 \end{cases}$$

$$T_n(s) = \begin{cases} \hat{S}_2(s) & \text{als } B_2(s) > 0 \\ 0 & \text{als } B_2(s) = 0. \end{cases}$$

Stel dat $S_m(s)$ en $T_n(s)$ geen gemeenschappelijke sprongtijden hebben, dan geldt:

$$\frac{1}{2}\int_0^\infty S_m(s)dT_n(s) = \frac{1}{2}[S_m(s)T_n(s)]_0^\infty - \frac{1}{2}\int_0^\infty T_n(s-)dS_m(s) =$$

$$= -\frac{1}{2} - \frac{1}{2}\int_0^\infty T_n(s-)dS_m(s).$$

Als de grootste waarneming in beide steekproeven ongecensureerd zijn, dan volgt dat $W = \frac{1}{2} + \frac{1}{2}Z_2(\infty)$, anders kan het verschil vrij groot zijn. We zullen $Z_2(\infty)$ aanduiden als de *gemodificeerde toetsingsgrootheid van Efron.*

In het vervolg zullen we eigenschappen van $Z_2(\infty)$ gaan onderzoeken. Voor zover ons bekend komen in de literatuur alleen heuristische en schetsmatige bewijzen voor van de eigenschappen van W. De eigenschappen van $Z_2(\infty)$ zullen in het algemeen niet gelden voor 2W - 1 maar zouden van geval tot geval apart onderzocht moeten worden. In paragraaf 3.4. gaan we de eigenschappen van de gemodificeerde produkt-limiet schatter en van de empirische cumulatieve uitvalsfunctie behandelen hetgeen van belang is voor eigenschappen van $Z_2(\infty)$.

3.3. DE VARIANTIES VAN DE DRIE TOETSINGSGROOTHEDEN

In 3.2. hebben we gezien dat

$$Z(t) = \int_0^t K(s)J(s)\left(\frac{dN_2(s)}{B_2(s)} - \frac{dN_1(s)}{B_1(s)}\right).$$

Onder H_0 geldt dat $\lambda_1(s) = \lambda_2(s) = \lambda(s)$ zodat

$$\langle Z,Z\rangle(t) = \int_0^t \frac{K^2(s)J(s)}{B_2(s)}\,\lambda(s)ds + \int_0^t \frac{K^2(s)J(s)}{B_1(s)}\,\lambda(s)ds =$$

$$= \int_0^t K^2(s)J(s)\,\frac{B_1(s)+B_2(s)}{B_1(s)B_2(s)}\,\lambda(s)ds.$$

In 3.2. hebben we gedefinieerd

(i) $K_1(t) = B_1(t)B_2(t)$ (Gehan)

(ii) $K_2(t) = \hat{S}_1(t-)\hat{S}_2(t-)$ (gemodificeerde Efron)

(iii) $K_3(t) = \dfrac{B_1(t)B_2(t)}{B_1(t)+B_2(t)} \cdot J(t)$ (Cox)

Verder

$$Z_i(t) = \int_0^t K_i(t)J(s)\left(\frac{dN_2(s)}{B_2(s)} - \frac{dN_1(s)}{B_1(s)}\right) \qquad i = 1,2,3.$$

Omdat $\{K^2(s)J(s)/B_1(s)B_2(s)\}$ een linkskontinu proces is, geldt dat

$$\int_0^t K^2(s)J(s)\ \frac{B_1(s)+B_2(s)}{B_1(s)B_2(s)}\ \frac{d(M_1(s)+M_2(s))}{B_1(s)+B_2(s)}$$

een kwadratisch integreerbare martingaal is.
Onder H_0 en als $\{K(s)\}$ begrensd is (in dat geval kunnen we stelling 2.2.17. toepassen omdat $\{B_1(s)\}$ en $\{B_2(s)\}$ begrensd zijn), kunnen we deze stochastische integraal schrijven als het verschil van twee Lebesgue-Stieltjes integralen

$$\int_0^t K^2(s)J(s)\ \frac{d(N_1(s)+N_2(s))}{B_1(s)B_2(s)} - \int_0^t K^2(s)J(s)\ \frac{B_1(s)+B_2(s)}{B_1(s)B_2(s)}\ \lambda(s)ds =$$

$$= \int_0^t K^2(s)J(s)\ \frac{d(N_1(s)+N_2(s))}{B_1(s)B_2(s)} - \langle Z,Z\rangle(t).$$

Onder H_0 is $E(Z(\infty)) = 0$ zodat wij $\mathrm{var}(Z(\infty)) = E(\langle Z,Z\rangle(\infty))$ zuiver kunnen schatten door

$$V(\infty) = \int_0^\infty \frac{K^2(s)J(s)}{B_1(s)B_2(s)}\ d(N_1(s)+N_2(s)).$$

Nu zijn $\{K_1(s)\}$, $\{K_2(s)\}$ en $\{K_3(s)\}$ begrensd dus stelling 2.2.17. is toepasbaar voor die gevallen onder H_0.
Als zuivere schatter voor de variantie van $Z_i(\infty)$ voor $i = 1,2,3$ hebben we achtereenvolgens:

$$V_1(\infty) = \int_0^\infty B_1(s)B_2(s)J(s)d(N_1(s)+N_2(s)) \qquad \text{(Gehan)}$$

$$V_2(\infty) = \int_0^\infty \frac{\hat{S}_1^2(s-)\hat{S}_2^2(s-)J(s)d(N_1(s)+N_2(s))}{B_1(s)B_2(s)} \qquad \text{(gemodificeerde Efron)}$$

$$V_3(\infty) = \int_0^\infty \frac{B_1(s)B_2(s)J(s)d(N_1(s)+N_2(s))}{(B_1(s)+B_2(s))^2} \qquad \text{(Cox)}$$

In definitie 1.5.4. hebben we voorgesteld dat de variantie van $U(0)$ te schatten was met $-D(0) = \sum_{i=1}^s \Pi_i(1-\Pi_i)$.

Merk op dat

$$V_3(\infty) = \int_0^\infty \frac{B_2(s)}{(B_1(s)+B_2(s))}\left(1 - \frac{B_2(s)}{B_1(s)+B_2(s)}\right)J(s)\,d(N_1(s)+N_2(s)) = -D(0).$$

$V_1(\infty)$ is niet dezelfde variantie als in lemma 1.2.4. omdat in dat lemma de voorwaardelijke variantie van $Z_1(\infty)$ is berekend. Merk op dat $V_1(\infty)$ makkelijker te berekenen is dan de voorwaardelijke variantie.

Onder bepaalde voorwaarden is het mogelijk om de asymptotische equivalentie van $\frac{1}{4}V_2(\infty)$ en de in lemma 1.4.2. genoemde variantie van W te bewijzen.

3.4. DE PRODUKT-LIMIET SCHATTER EN DE EMPIRISCHE CUMULATIEVE UITVALSFUNCTIE

Stel dat een stochastische grootheid een verdelingsfunctie F heeft met dichtheid f.
De uitvalsfunctie wordt dan gedefinieerd als $\lambda(t) = \frac{f(t)}{1 - F(t)}$.

DEFINITIE 3.4.1. De *cumulatieve uitvalsfunctie* $\gamma(t)$ definieert men als

$$\gamma(t) = \int_0^t \lambda(s)ds.$$

EIGENSCHAP 3.4.2.

$$\gamma(t) = \int_0^t \frac{f(s)}{1 - F(s)}\,ds = -\log(1-F(t)) = -\log S(t)$$

waarbij S(t) de overlevingsfunctie voorstelt.

Stel $X_1, X_2, \ldots, X_n$ zijn o.o. identiek verdeelde stochastische grootheden die niet-negatief zijn en uitvalsfunctie $\lambda(t)$ hebben.
Stel $U_1, U_2, \ldots, U_n$ zijn niet-negatief (met kans 1) en er bestaan σ-algebra's $\{\mathcal{F}_t\}$ zodanig dat het telproces $\{N(t)\} = \{\#\{i: \tilde{X}_i \leq t \text{ en } \Delta_i = 1\}\}$ met $\tilde{X}_i = X_i \wedge U_i$ en $\Delta_i = I_{\{\tilde{X}_i = X_i\}}$, intensiteitsproces $\{\lambda(t)B(t)\}$ t.o.v. $\{\mathcal{F}_t\}$ heeft met $B(t) = \#\{i: \tilde{X}_i \geq t\}$.
Definieer $M(t) = N(t) - \int_0^t \lambda(s)B(s)ds$.
De gemodificeerde produkt-limiet schatter van $S(t) = P(X_i > t)$ is gedefinieerd als

$$\hat{S}(t) = \prod_{\{s\leq t:\exists i \text{ z.d.d. } \tilde{X}_i = s, \Delta_i = 1\}} \frac{\#\{j: \tilde{X}_j \geq s\} - 1}{\#\{j: \tilde{X}_j \geq s\}} =$$

$$= \prod_{\{s\leq t:N(s)=N(s-)+1\}} (1-B^{-1}(s)).$$

De eigenschappen van de (gemodificeerde) produkt-limiet schatter zijn het eenvoudigst af te leiden uit die van de empirische cumulatieve uitvalsfunctie, zoals deze is geïntroduceerd door NELSON [21].

DEFINITIE 3.4.3. De *empirische cumulatieve uitvalsfunctie* $\hat{G}(t)$ wordt gedefinieerd als

$$\hat{G}(t) = \int_0^t B^{-1}(s)dN(s).$$

STELLING 3.4.4. *Met kans* 1 *geldt voor alle* t,τ *die voldoen aan* $t \le \tau$ *en* $B(\tau) > 0$ *dat*

$$0 \le -\log(\hat{S}(t)) - \hat{G}(t) < \frac{(m-B(t))}{mB(t)} \le \frac{(m-B(\tau))}{mB(\tau)} = \frac{1}{m} \cdot \frac{1-B(\tau)/m}{B(\tau)/m}.$$

BEWIJS. Zie BRESLOW & CROWLEY [6] p.445-446. Het bewijs volgt uit de ongelijkheid $0 < -\log(1-(x+1)^{-1}) - (x+1)^{-1} < (x(x+1))^{-1} \forall x > 0$.

OPMERKING 3.4.5. Als er een $\tau > 0$ bestaat zodanig dat $\frac{B(\tau)}{m} \xrightarrow{P} c > 0$ als $m \to \infty$ dan konvergeert $m^{\frac{1}{2}}(-\log(\hat{S}(t)-\hat{G}(t))$ uniform in t met $t \in [0,\tau]$ in waarschijnlijkheid naar nul. Men zou dit kunnen gebruiken om de equivalentie van de asymptotische verdelingen te bewijzen.

Definieer $J(t) = I_{\{B(t)>0\}}$ en als $B(t) = 0$ dan $\frac{J(t)}{B(t)} = 0$.
$\{J(t)B^{-1}(t)\}$ is een begrensd linkskontinu proces horend bij $\{F_t\}$, dus $\{J(t)B^{-1}(t)\}$ is voorspelbaar.
De integraal

$$\int_0^t J(s)B^{-1}(s)dM(s) = \int_0^t J(s)B^{-1}(s)dN(s) - \int_0^t J(s)\lambda(s)ds$$

is gedefinieerd als stochastische integraal en als Lebesgue-Stieltjes integraal (zie stelling 2.2.17.).
$\int_0^t J(s)B^{-1}(s)dM(s)$ is een kwadratisch integreerbare martingaal waarvoor geldt:

$$\int_0^t J(s)B^{-1}(s)dM(s) = \int_0^t B^{-1}(s)dN(s) - \int_0^t J(s)\lambda(s)ds =$$

$$= \hat{G}(t) - \gamma(t \wedge T)$$

$$= \hat{G}(t \wedge T) - \gamma(t \wedge T).$$

waarbij T gedefinieerd is door $T = \inf\{t : B(t)=0\}$ dus T is een stoptijd!

Voor iedere stoptijd T' geldt $E(\hat{G}(T' \wedge T)) = E(\gamma(T' \wedge T))$ (vanwege martingaal-eigenschap (zie stelling 2.2.12.) zodat in zekere zin $\hat{G}$ een zuivere schatter is van γ.

Wegens gevolg 2.1.33.(i) geldt dat

$$<\hat{G}(t)-\gamma(t\wedge T),\hat{G}(t)-\gamma(t\wedge T)> = \int_0^t \left(\frac{J(s)}{B(s)}\right)^2 \lambda(s)B(s)ds =$$

$$= \int_0^t \frac{J(s)}{B(s)} \lambda(s)ds.$$

Per definitie is $(\hat{G}(t)-\gamma(t\wedge T))^2 - <\hat{G}(t)-\gamma(t\wedge T),\hat{G}(t)-\gamma(t\wedge T)>$ een martingaal dus

$$E(\hat{G}(t)-\gamma(t\wedge T))^2 = \int_0^t E\left(\frac{J(s)}{B(s)}\right) \lambda(s)ds \qquad \text{(zie stelling 2.1.15.).}$$

Nu is $\left(\frac{J(s)}{B(s)}\right)^2$ begrensd en voorspelbaar, dus volgens stelling 2.2.17. is de stochastische integraal $\int_0^t \left(\frac{J(s)}{B(s)}\right)^2 dM(s)$ gelijk aan de Lebesgue-Stieltjes integraal $\int_0^t \left(\frac{J(s)}{B(s)}\right)^2 dN(s) - \int_0^t \frac{J(s)}{B(s)} \lambda(s)ds$, zodat $\int_0^t E\left(\frac{J(s)}{B(s)}\right) \lambda(s)ds$ zuiver geschat kan worden door $\int_0^t \frac{J(s)}{B^2(s)} dN(s)$.

We gaan nu de asymptotische verdeling van de empirische cumulatieve uitvalsfunctie afleiden in een aantal gevallen.

Stel dat men voor iedere n de volgende stochastische vektoren kan waarnemen $(\tilde{X}_{n,i},\Delta_{n,i})$ voor $i = 1,2,\ldots,n$, waarbij $\tilde{X}_{n,i} = X_{n,i} \wedge U_{n,i}$, waarin $X_{n,i}$ en $U_{n,i}$ aan dezelfde eisen voldoen als in het begin van de paragraaf.

Voor iedere n hebben we dan een drietal processen nl. $\{N_n(t)\}$, $\{B_n(t)\}$ en $\{J_n(t)\}$ die gedefinieerd zijn door

$$N_n(t) = \#\{i: \tilde{X}_{n,i} \le t \text{ én } \Delta_{n,i} = 1\}$$
$$B_n(t) = \#\{i: \tilde{X}_{n,i} \ge t\}$$
$$J_n(t) = I_{\{B_n(t)>0\}} \qquad \text{voor alle } t \in [0,\infty).$$

Verder moet gelden dat er voor iedere n een familie σ-algebra's $\{F_{n,t}\}$ bestaat zodanig dat het telproces $\{N_n(t)\}$ intensiteitsproces $\{\lambda(t)B_n(t)\}$ t.o.v. $\{F_{n,t}\}$ heeft dwz. $\{N_n(t) - \int_0^t\lambda(s)B_n(s)ds\}$ een $\{F_{n,t}\}$-martingaal is.

<u>OPMERKING 3.4.6.</u> De functie $\lambda(t)$ is voor elke steekproef dezelfde.

De empirische cumulatieve uitvalsfunctie van de n-de steekproef definiëren we door

$$\hat{G}_n(t) = \int_0^t B_n^{-1}(s)dN_n(s).$$

We definiëren
$$G_n(t) = \gamma(t\wedge T_n) = \int_0^t J_n(s)\lambda(s)ds \qquad \text{waarbij } T_n = \inf\{t\colon B_n(t) = 0\}.$$

Stel I is een willekeurig gesloten interval bevat in $[0,\tau)$, waarin $\tau \leq \infty$ zodanig dat $F(\tau) = 1$ en $F(t) < 1$ is voor alle $t < \tau$, zie paragraaf 3.1. Voor het gemak nemen we $I = [0,1]$ (eis $\tau > 1$).
We gaan de asymptotische verdeling van $n^{\frac{1}{2}}(\hat{G}_n(t)-G_n(t))$ bepalen voor $t \in I$.

Om stelling 2.3.1. toe te passen moet aan de volgende voorwaarden voldaan zijn.

(a) Er bestaat een functie g op $[0,1]$ met $\int_0^1 g^2(t)dt < \infty$ zodanig dat
$$\int_0^t n\frac{J_n(s)}{B_n(s)}\lambda(s)ds \overset{P}{\to} \int_0^t g^2(s)ds \qquad \text{voor alle } t \in [0,1].$$

(b) Voor alle $\varepsilon > 0$ geldt
$$nE\int_0^1 B_n^{-2}(t)I_{\{B_n(t)<n^{\frac{1}{2}}\varepsilon^{-1}\}}dN_n(t) \to 0 \qquad \text{als } n \to \infty.$$

(c) Voor iedere n bestaat er een niet-dalend proces $\{Z_n(t)\}$ en een stochastische grootheid C_n zodanig dat

(i) $|\lambda(t)J_n(t)B_n^{-1}(t) - \lambda(s)J_n(s)B_n^{-1}(s)| \leq Z_n(t) - Z_n(s)$

voor alle $0 \leq s \leq t \leq 1$ en $E(Z_n(1)-Z_n(0)) < \infty$.

(ii) $\lambda(t)J_n(t)B_n^{-1}(t) \leq C_n$

voor alle $0 \leq t \leq 1$ en $E(C_nN_n(1)) < \infty$ en $E(C_n\int_0^1\lambda(t)B_n(t)dt) < \infty$.

STELLING 3.4.7. *Als aan de voorwaarden* (a), (b) *en* (c) *is voldaan dan geldt*
$$n^{\frac{1}{2}}(\hat{G}_n(t)-G_n(t)) \overset{zw}{\to} \int_0^t g(s)dW(s)$$
(in D[0,1]*, waarin* W *het Wiener proces is).*

BEWIJS. Zie stelling 2.3.1.

STELLING 3.4.8. *Als* $\lambda(t)$ *van begrensde variatie op* [0,1] *is, is aan voorwaarde* (c) *voldaan.*

BEWIJS.

(i) $$|\lambda(t)J_n(t)B_n^{-1}(t) - \lambda(s)J_n(t)B_n^{-1}(t) + \lambda(s)J_n(t)B_n^{-1}(t) - \lambda(s)J_n(s)B_n^{-1}(s)| \le$$

$$\le |\lambda(t) - \lambda(s)| + |\lambda(s)|\,|J_n(t)B_n^{-1}(t) - J_n(s)B_n^{-1}(s)| \le$$

$$\le |\lambda(t) - \lambda(s)| + \sup_{s\in[0,1]} |\lambda(s)|\,|-B_n(t) + B_n(s)| \le$$

$$\le \beta(t) - \beta(s) + \sup_{s\in[0,1]} |\lambda(s)|(-B_n(t)+B_n(s)).$$

Immers $\lambda(t)$ is van begrensde variatie dus er bestaat een niet-dalende functie $\beta(t)$ zodanig dat $|\lambda(t) - \lambda(s)| \le \beta(t) - \beta(s)$ voor alle $0 \le s \le \le t \le 1$. Verder maakt het proces $\{\lambda(t)J_n(t)B_n^{-1}(t)\}$ sprongen ter grootte van hoogstens $\sup_{s\in[0,1]} |\lambda(s)|$ op de sprongtijden van het proces $\{B_n(t)\}$.
Kies $Z_n(t) = \beta(t) - \sup_{s\in[0,1]} |\lambda(s)|.B_n(t)$ voor alle $t \in [0,1]$.

(ii) $\lambda(t)J_n(t)B_n^{-1}(t)$ is begrensd en $E(N_n(1)) = E(\int_0^1\lambda(t)B_n(t)dt) < n$. □

TOEPASSING 3.4.9.

(i) Stel $U_1,U_2,\ldots,U_n,\ldots$ zijn o.o. identiek verdeelde stochastische grootheden die de waarde ∞ kunnen aannemen. Zij $H(t) = P(U_1<t)$.
Verder zijn $X_1,X_2,\ldots,X_n,\ldots$ o.o. identiek verdeelde stochastiche grootheden met kontinue verdelingsfunctie F en uitvalsfunctie $\lambda(t)$.
De rijen $\{X_i\}$, $\{U_i\}$ zijn onafhankelijk.
We gaan nu na of $\{n^{\frac{1}{2}}(\hat{G}_n(t) - \gamma(t))\}$ zwak konvergeert naar een stochastisch proces hetgeen een sterkere uitspraak is dan die van stelling 3.4.7. We verifiëren eerst de voorwaarden (a), (b) en (c) van stelling 3.4.7.
Stel dat $\lambda(t)$ op $[0,1]$ van begrensde variatie is (men zou in deze situatie zelfs kunnen volstaan met begrensdheid van λ te eisen). Hieruit volgt dat $F(1) < 1$ omdat $-\log(1-F(1)) = \int_0^1\lambda(s)ds < \infty$. Stel verder dat $H(1) < 1$.
We definiëren hier $U_{ni} = U_i$ en $X_{ni} = X_i$ voor alle $i \le n$.
Toepassing van de stelling van Glivenko-Cantelli (zie LOÈVE [17]p.20-21) geeft dat $\frac{B_n(s)}{n} \to (1-H(s))(1-F(s))$ uniform in $s \in [0,1]$, met kans 1 als $n \to \infty$.
Bovendien geldt dat $\frac{\lambda(s)}{(1-H(s))(1-F(s))}$ begrensd is op $[0,1]$ omdat $F(1) < 1$, $H(1) < 1$ en $\lambda(t)$ van begrensde variatie is.

Dus

$$\int_0^1 \frac{\lambda(s)}{(1-H(s))(1-F(s))}\, ds < \infty$$

en

$$\int_0^t n\, \frac{J_n(s)}{B_n(s)}\, \lambda(s)ds \overset{P}{\to} \int_0^t g^2(s)ds \qquad \forall t \in [0,1]$$

waarbij $g^2(t) = \dfrac{\lambda(s)}{(1-H(t))(1-F(t))}$. Aan voorwaarde (a) is dus voldaan.

Nu is $N_n(t) - \int_0^t \lambda(s)B_n(s)ds$ een martingaal t.o.v. $\{\mathcal{F}_{nt}\}$ en $B_n^{-2}(t)J_n(t)I_{\{B_n(t)<n^{\frac{1}{2}}\varepsilon-1\}}$ linkskontinu, begrensd en horend bij $\{\mathcal{F}_{nt}\}$. Om voorwaarde (b) te verifiëren moet men bewijzen dat $\forall\varepsilon > 0$

$$nE\left(\int_0^1 J_n(t)B_n^{-1}(t)I_{\{B_n(t)<n^{\frac{1}{2}}\varepsilon-1\}}\lambda(t)dt\right) \to 0 \qquad \text{als } n \to \infty.$$

Door gebruikmaking van de stelling van Fubini, de ongelijkheid van Cauchy-Schwarz en het feit dat $B_n(t)$ niet-stijgend is, geldt:

$$E\left(\int_0^1 \frac{nJ_n(t)}{B_n(t)}\, I_{\{B_n(t)<n^{\frac{1}{2}}\varepsilon-1\}}\lambda(t)dt\right) = \int_0^1 E\left(\frac{nJ_n(t)}{B_n(t)}\, I_{\{B_n(t)<n^{\frac{1}{2}}\varepsilon-1\}}\right)\lambda(t)dt \le$$

$$\le \int_0^1 \left[E\left(\frac{n^2J_n(t)}{B_n^2(t)}\right)\right]^{\frac{1}{2}} \left[P(B_n(t)<n^{\frac{1}{2}}\varepsilon^{-1})\right]^{\frac{1}{2}}\lambda(t)dt \le$$

$$\le \left[P(B_n(1) < n^{\frac{1}{2}}\varepsilon^{-1})\right]^{\frac{1}{2}} \int_0^1 \left[E\left(\frac{6(n+1)(n+2)}{(B_n(t)+1)(B_n(t)+2)}\right)\right]^{\frac{1}{2}} \lambda(t)dt.$$

Zij X een stochastische grootheid met een binomiale verdeling met parameters n en p

$$E\left(\frac{(n+1)(n+2)}{(X+1)(X+2)}\right) = \sum_{x=0}^{n} \binom{n}{x} \frac{(n+1)(n+2)}{(x+1)(x+2)}\, p^x(1-p)^{n-x} =$$

$$= \frac{1}{p^2} \sum_{x=2}^{n+2} \binom{n+2}{x} p^x(1-p)^{n+2-x} \le$$

$$\le \frac{1}{p^2} \sum_{x=0}^{n+2} \binom{n+2}{x} p^x(1-p)^{n+2-x} = \frac{1}{p^2}\,.$$

Daar $B_n(t)$ binomiaal verdeeld is met parameters n en $(1-H(t))(1-F(t))$ kunnen we deze ongelijkheid ook gebruiken voor ons geval

$$\int_0^1 \left[E\left(\frac{6(n+1)(n+2)}{(B_n(t)+1)(B_n(t)+2)}\right)\right]^{\frac{1}{2}} \lambda(t)dt \le \int_0^1 \frac{\sqrt{6}\,\lambda(t)}{(1-H(t))(1-F(t))}\, dt < \infty$$

omdat $F(1) < 1$ én $H(1) < 1$.

Verder geldt dat $\frac{B_n(t)}{n} \xrightarrow{P} (1-H(t))(1-F(t))$ dus $P(B_n(1)<n^{\frac{1}{2}}\varepsilon^{-1}) \to 0$.

Voorwaarde (b) is dus vervuld omdat

$$[P(B_n(1)<n^{\frac{1}{2}}\varepsilon^{-1})]^{\frac{1}{2}} \int_0^1 \frac{\sqrt{6}\,\lambda(t)}{(1-H(t))(1-F(t))}\, dt \to 0.$$

Voorwaarde (c) is vervuld omdat $\lambda(t)$ van begrensde variatie is op $[0,1]$.

Dus $n^{\frac{1}{2}}(\hat{G}_n(t)-G_n(t)) \xrightarrow{zw} \int_0^t g(s)dW(s)$ op $[0,1]$.

Daar $P(J_n(1)=1) \to 1$ als $n \to \infty$ kunnen we $G_n(t)$ vervangen door $\gamma(t)$.

(ii) Stel $U_1,U_2,\ldots$ zijn o.o., laat $X_1,X_2,\ldots$ o.o. en identiek verdeeld zijn en de rijen $\{X_i\}$, $\{U_i\}$ zijn onafhankelijk.

Van de kontinue verdelingsfunctie F van X_1 eisen we dat $0 < F(1) < 1$.

De uitvalsfunctie $\lambda(x)$ is van begrensde variatie op $[0,1]$ (ook hier zou met begrensdheid volstaan kunnen worden).

Aan voorwaarde (c) van stelling 3.4.7. is weer voldaan (zie stelling 3.4.8.).

Stel dat $\frac{1}{n}\sum_{i=1}^n P(U_i<s) \to H(s)$ als $n \to \infty$ voor zekere functie H met $H(1) < 1$.

Met behulp van een type Glivenko-Cantelli stelling voor niet-identiek verdeelde stochastische grootheden (zie VAN ZUYLEN [29] gevolg 1.3.1., p.30-31) geldt

$$\sup_{s\in[0,1]} \left|\frac{B_n(s)}{n} - (1-H(s))(1-F(s))\right| \xrightarrow{P} 0 \qquad \text{als } n \to \infty.$$

Aan voorwaarde (a) is dus ook voldaan (zie (i)).

Uit (i) hebben we de ongelijkheid $\frac{n^2 J_n(t)}{B_n(t)} \le \frac{6(n+1)(n+2)}{(B_n(t)+1)(B_n(t)+2)}$.

Nu is $f(x) = \frac{1}{(x+1)(x+2)}$ een strikt convexe functie op $[0,\infty)$ en $B_n(t)$ de som van n onafhankelijke Bernoulli stochasten. Volgens HOEFFDING [15] (stelling 3,p.717-718) geldt dat $E\left(\frac{1}{(B_n(t)+1)(B_n(t)+2)}\right)$ maximaal is bij konstante $E(B_n(t))$ als $B_n(t)$ binomiaal verdeeld is met parameters n en p met $p = \frac{1}{n}\sum_{i=1}^n P(\tilde{X}_i \ge t)$.

Dus

$$E\left(\frac{1}{(B_n(t)+1)(B_n(t)+2)}\right) \le \frac{1}{p^2} \xrightarrow{P} \frac{1}{(1-H(t))^2(1-F(t))^2} \qquad \text{voor } n \to \infty.$$

Analoog aan (i) geldt $P(B_n(1)<n^{\frac{1}{2}}\varepsilon^{-1}) \to 0$ zodat voorwaarde (b) vervuld is (zie (i)).

Hieruit volgt dat $n^{\frac{1}{2}}(\hat{G}_n(t)-G_n(t)) \xrightarrow{zw} \int_0^t g(s)dW(s)$ op $[0,1]$ met

$g^2(s) = \dfrac{\lambda(s)}{(1-H(s))(1-F(s))}$.

Bovendien geldt dat $n^{\frac{1}{2}}(\hat{G}_n(t)-\gamma(t)) \xrightarrow{zw} \int_0^t g(s)dW(s)$ op $[0,1]$ omdat $P(J_n(1)=1) \to 1$.

In toepassing 3.4.9. hebben we onder bepaalde voorwaarden zwakke konvergentie kunnen bewijzen van het proces $\{n^{\frac{1}{2}}(\hat{G}(t)-\gamma(t))\}$ op $[0,1]$.
Op grond van stelling 3.4.4. hebben we de volgende ongelijkheid

$$0 < -\log \hat{S}_n(t) - \hat{G}_n(t) < n^{-1}\,\frac{1-B_n(1)/n}{B_n(1)/n} \qquad \text{als } B_n(1) > 0 \text{ en } t \in [0,1].$$

Stel dat $\dfrac{B_n(1)}{n} \xrightarrow{P} c > 0$ als $n \to \infty$ hetgeen het geval was in toepassing 3.4.9.
In kombinatie met de vorige ongelijkheid volgt door Taylor ontwikkeling dat

$$\sup_{t\in[0,1]} \left| n^{\frac{1}{2}}(\hat{S}_n(t)-S(t)) - n^{\frac{1}{2}}(-S(t))[\hat{G}_n(t) - \gamma(t)] \right| \xrightarrow{P} 0, \qquad \text{als } n \to \infty,$$

(zie BRESLOW & CROWLEY [6] p.448). Indien $n^{\frac{1}{2}}(\hat{G}_n(t)-\gamma(t))$ zwak konvergeert dan volgt dat

$$n^{\frac{1}{2}}(\hat{S}_n(t)-S(t)) \xrightarrow{zw} (1-F(t)) \int_0^t g(s)dW(s), \qquad \text{als } t \in [0,1].$$

In bepaalde gevallen kunnen wij dus de asymptotische verdeling van de produkt-limiet schatter bepalen.

OPMERKING 3.4.18. Beschouw de situatie van toepassing 3.4.9.(ii). Als $\tau \in (0,\infty]$ zodanig dat $(1-H(t))(1-F(t)) > 0$ voor alle $t \in [0,\tau)$ en $\lambda(t)$ van begrensde variatie is op ieder gesloten deelinterval van $[0,\tau)$, heeft men zwakke konvergentie van $n^{\frac{1}{2}}(\hat{S}_n(t)-S(t))$ op $[0,\sigma]$ voor ieder $\sigma < \tau$. Toepassing van stelling 2.3.3. met $A = [0,\tau)$ en kompakte deelintervallen K zodanig dat $K \subset [0,\sigma] \subset [0,\tau)$ levert dan zwakke konvergentie op $[0,\tau)$.

3.5. ASYMPTOTISCHE RESULTATEN

In deze paragraaf zullen wij verschillende keren stelling 2.3.1. gebruiken. We zullen zien dat als $\lambda(t)$ van begrensde variatie is op een gesloten interval $[0,\tau]$ met $F(\tau) < 1$, aan voorwaarde (c) van stelling 2.3.1. voldaan

is.
Als men het bewijs van stelling 2.3.1. nagaat dan blijkt dat voorwaarde (c) in de gevallen die wij beschouwen, verzwakt kan worden tot de voorwaarde $F(x)$ heeft een begrensde afgeleide op $[0,\tau]$ en $F(\tau) < 1$, of equivalent, $\lambda(t)$ is begrensd op $[0,\tau]$ en $\tau < \infty$.

In deze paragraaf beperken we ons tot het volgende geval:
Stel $\{X_i\}$, $\{U_i\}$, $\{Y_j\}$, $\{V_j\}$ zijn onafhankelijke rijen van o.o. identiek verdeelde stochastische grootheden.
We werken nu verder onder de nulhypothese H_0: de verdelingsfunctie F van X_1 is gelijk aan de verdelingsfunctie G van Y_1.
Stel dat $F(x) < 1$ voor alle $x \in [0,\tau)$ en dat F een uitvalsfunctie $\lambda(t)$ heeft, die op gesloten deelintervallen van begrensde variatie is. Verder heeft U_i voor alle i verdelingsfunctie J en V_j heeft voor alle j verdelingsfunctie K (die in tegenstelling tot F niet noodzakelijk kontinu zijn).

Definieer $\sigma = \inf\{s: J(s) = 1$ óf $K(s) = 1$ óf $F(s) = 1\}$.
Dan geldt dus dat $Z_i(\infty) = Z_i(\sigma)$ voor $i = 1,2,3$ omdat tenminste één van de processen $\{B_1(t)\}$, $\{B_2(t)\}$ nul is op (σ,∞).

In het volgende onderscheiden we vier gevallen (alles onder H_0)

A: $J(\sigma-) < 1$, $K(\sigma-) < 1$ en $F(\sigma) < 1$ (dus $\sigma < \infty$)

B: $J(\sigma-) < 1$ òf $K(\sigma-) < 1$ en $F(\sigma) < 1$ (dus $\sigma < \infty$)

C: $J(\sigma-) = 1$, $K(\sigma-) = 1$ en $F(\sigma) < 1$ (dus $\sigma < \infty$)

D: $F(\sigma) = 1$.

Stel dat

$$Z(t) = \nu^{-\frac{1}{2}}\left[\int_0^t H_2(s)dN_2(s) - \int_0^t H_1(s)dN_1(s)\right]$$

onder H_0 een martingaal is waarbij $\nu = m + n$ ($H_1(s)$ en $H_2(s)$ hebben betrekking op onze toetsen). Ter vereenvoudiging van de notatie wordt de index ν steeds weggelaten (bijvoorbeeld bij H_i en N_i).

LEMMA 3.5.1. *Stelling* 2.3.1. *is geldig in de gevallen* A, B *en* C *als*

(i) $$\int_0^t \nu^{-1}H_i^2(s)B_i(s)\lambda(s)ds \xrightarrow{P} \int_0^t g_i^2(s)ds < \infty \qquad \forall t \in [0,\sigma] \text{ als } \nu \to \infty,$$ voor $i = 1,2$.

(ii) $H_i(s)$ *is een trapfunctie met sprongtijden bevat in die van* $B_1(s) + B_2(s)$, *uniform begrensd in* ν *en* s *door een konstante*.

BEWIJS. Voorwaarde (ii) en λ van begrensde variatie op $[0,\sigma]$ impliceren dat aan voorwaarde (c) is voldaan (bewijs analoog aan het bewijs van stelling 3.4.8.).

Voorwaarde (a) van stelling 2.3.1. is gelijk aan voorwaarde (i). We moeten dus alleen nog aantonen dat

$$E\left\{\int_0^\sigma \nu^{-1}H_i^2(s)I_{\{|\nu^{-\frac{1}{2}}H_i(s)|>\varepsilon\}}dN_i(s)\right\} \to 0 \quad \text{als } \nu\to\infty, \text{ voor } i=1,2.$$

Voor alle i, s en ν geldt dat $H_i(s) \le c$, dus geldt $I_{\{|\nu^{-\frac{1}{2}}H_i(s)|>\varepsilon\}} \equiv 0$ voor alle $\nu \ge \left(\frac{c}{\varepsilon}\right)^2$. □

STELLING 3.5.2. *Indien* $\frac{m}{\nu} \to \rho$ *met* $0 < \rho < 1$ *dan is stelling* 2.3.1. *geldig in de gevallen* A, B *en* C *voor de toetsingsgrootheid van Gehan, in de gevallen* A *en* B *voor de toetsingsgrootheid van Cox en in het geval* A *voor de gemodificeerde toetsingsgrootheid van Efron.*

BEWIJS. Voor de geschikt genormeerde toetsingsgrootheid van Gehan hebben we $H_1(s) = \frac{B_2(s)}{\nu}$ en $H_2(s) = \frac{B_1(s)}{\nu}$ met $\nu = m + n$ (vergelijk bewijs stelling 3.2.4.).
Dus $H_1(s), H_2(s) < 1$ voor alle ν en s.
Verder

$$\nu^{-1}H_1^2(s)B_1(s)\lambda(s) = \left(\frac{B_2(s)}{\nu}\right)^2\left(\frac{B_1(s)}{\nu}\right)\lambda(s).$$

Vanwege de stelling van Glivenko-Cantelli konvergeert deze uitdrukking in waarschijnlijkheid uniform in $s \in [0,\sigma]$ naar $(1-\rho)^2\rho(1-\tilde{G}(s-))^2(1-\tilde{F}(s-))\lambda(s)$ als $\nu \to \infty$.

Omdat $\int_0^\sigma \lambda(s)ds < \infty$ geldt

$$\int_0^t \nu^{-1}H_1^2(s)B_1(s)\lambda(s)ds \overset{P}{\to} (1-\rho)^2\rho \int_0^t (1-\tilde{G}(s-))^2(1-\tilde{F}(s-))\lambda(s)ds =$$

$$= \int_0^t g_1^2(s)ds < \infty, \qquad \text{als } \nu \to \infty.$$

Analoog voor $H_2(s)$.
Voor de gevallen A, B en C is aan de voorwaarden van lemma 3.5.1. voldaan.

Voor de geschikt genormeerde toetsingsgrootheid van Cox hebben we

$$H_1(s) = \frac{B_2(s)}{B_1(s)+B_2(s)} I_{\{B_1(s)>0,B_2(s)>0\}} \text{ en } H_2(s) = \frac{B_1(s)}{B_1(s)+B_2(s)} I_{\{B_1(s)>0,B_2(s)>0\}}$$

dus $H_i(s) \le 1$ voor alle s en i = 1,2.
Wegens de stelling van Glivenko-Cantelli geldt $\frac{B_2(s)}{\nu} \overset{P}{\to} (1-\rho)(1-\tilde{G}(s-))$ uniform in $s \in [0,\sigma]$ als $\nu \to \infty$ en $\frac{B_1(s)+B_2(s)}{\nu} \overset{P}{\to} (1-\tilde{H}(s-))$ uniform in $s \in [0,\sigma]$ als $\nu \to \infty$ ($\tilde{H}(s) = \rho\tilde{F}(s) + (1-\rho)\tilde{G}(s) = 1 - (1-F(s))(1-L(s))$ met $L(s) = \rho J(s) + (1-\rho)K(s)$).

Voor de gevallen A en B geldt $\int_0^\sigma \lambda(s)ds < \infty$ en $(1-\tilde{H}(\sigma-)) > 0$ dus

$$\int_0^t \nu^{-1}\left(\frac{B_2(s)}{B_1(s)+B_2(s)}\right)^2 B_1(s)\lambda(s)ds \xrightarrow{P} \rho(1-\rho)^2 \int_0^t \frac{(1-\tilde{F}(s-))(1-\tilde{G}(s-))^2}{(1-\tilde{H}(s-))^2}\lambda(s)ds$$

voor alle $t \in [0,\sigma]$ en $\nu \to \infty$.

Analoog voor $H_2(s)$.

Lemma 3.5.1. is toepasbaar in de gevallen A en B.

Voor de geschikt genormeerde gemodificeerde toetsingsgrootheid van Efron hebben we $H_1(s) = \dfrac{\hat{S}_1(s-)\hat{S}_2(s-)}{B_1(s)}\nu$ en $H_2(s) = \dfrac{\hat{S}_1(s-)\hat{S}_2(s-)}{B_2(s)}\nu$.

In opmerking 3.4.10. hebben we zwakke konvergentie bewezen voor $m^{\frac{1}{2}}(\hat{S}_1(s)-S(s))$ als $m \to \infty$ en $n^{\frac{1}{2}}(\hat{S}_2(s)-S(s))$ als $n \to \infty$ op $[0,\sigma]$ waarin $S(s) = 1 - F(s)$. Dit geeft uniforme konvergentie in waarschijnlijkheid van $\nu^{-1}H_1^2(s)B_1(s)\lambda(s)$ naar $\rho^{-1}\dfrac{(1-F(s))^4}{(1-\tilde{F}(s-))}\lambda(s)$ voor $s \in [0,\sigma]$ en iets dergelijks voor $\nu^{-1}H_2^2(s)B_2(s)\lambda(s)$.

Dus lemma 3.5.1. is toepasbaar voor het geval A voor de gemodificeerde toetsingsgrootheid van Efron. $\square$

Voor iedere toetsingsgrootheid gaan we nu zwakke konvergentie na voor de overige gevallen.

Uit stelling 3.5.2. volgt dat zwakke konvergentie ook bewezen is op $[0,\theta]$ voor alle $\theta < \sigma$ want als $\theta < \sigma$ dan geldt dat $J(\theta-) < 1$, $K(\theta-) < 1$ en $F(\theta) < 1$.

Toepassing van stelling 2.3.3. geeft zwakke konvergentie op $[0,\sigma)$.

Uitbreiding tot $[0,\sigma]$ is alleen mogelijk als $\int_0^\sigma g_i^2(s)ds < \infty$, $i = 1,2$.

Voor de toetsingsgrootheid van Gehan geldt dat $\int_0^\sigma (g_1^2(s)+g_2^2(s))ds =$
$= (1-\rho)\rho \int_0^\sigma (1-\tilde{F}(s-))(1-\tilde{G}(s-))(1-L(s-))dF(s) \leq \rho(1-\rho)F(\sigma) \leq 1$ met
$L(s) = \rho J(s) + (1-\rho)K(s)$.

Voor de toetsingsgrootheid van Cox geldt dat

$$\frac{(1-J(s-))(1-K(s-))}{1-L(s-)} = \frac{1}{\rho(1-K(s-))^{-1} + (1-\rho)(1-J(s-))^{-1}} \leq 1 \quad \text{op } [0,\sigma)$$

(immers op $[0,\sigma)$ geldt dat $K(s),J(s) > 0$).

Dus

$$\int_0^\sigma (g_1^2(s)+g_2^2(s))ds = \rho(1-\rho)\int_0^\sigma \frac{(1-J(s-))(1-K(s-))}{(1-L(s-))}dF(s) \leq$$

$$\leq \rho(1-\rho)F(\sigma) \leq 1.$$

Voor de gemodificeerde toetsingsgrootheid van Efron geldt dat

$$\int_0^\sigma (g_1^2(s)+g_2^2(s))ds = \rho^{-1}\int_0^\sigma \frac{(1-F(s))^4}{(1-\tilde{F}(s-))}\lambda(s)ds + (1-\rho)^{-1}\int_0^\sigma \frac{(1-F(s))^4}{(1-\tilde{G}(s-))}\lambda(s)ds =$$

$$= \rho^{-1}(1-\rho)^{-1}\int_0^\sigma \frac{(1-F(s))^2(1-L(s-))}{(1-J(s-))(1-K(s-))}dF(s).$$

Er moet gelden dat $\int_0^\sigma (g_1^2(s)+g_2^2(s))ds < \infty$ dus moeten we voor de gemodificeerde toetsingsgrootheid van Efron veronderstellen dat

$$\int_0^\sigma \frac{(1-F(s))^2(1-L(s-))}{(1-J(s-))(1-K(s-))}dF(s) < \infty.$$

Dat dit laatste niet altijd waar is, blijkt uit het volgende voorbeeld.

VOORBEELD 3.5.3. Stel $1 - J(s) = 1 - K(s) = (1-F(s))^3$ dan is bij de gemodificeerde toetsingsgrootheid van Efron

$$\rho(1-\rho)\sum_{i=1}^{2}\int_0^\infty g_i^2(s)ds = \int_0^\infty (1-F(s))^{-1}dF(s) = [-\log(1-F(s))]_0^\infty = \infty,$$
$$i = 1,2.$$

Bij de toetsingsgrootheden van Gehan en Cox geldt voor alle (sub-) verdelingsfuncties J, K en F dat $\int_0^\infty g_i^2(s)ds < \infty$, $i = 1,2$.

We wensen de asymptotische verdeling van $Z(\sigma)$ te kennen. Dit gaan we als volgt aanpakken: Zoek een niet-negatieve functie q zodanig dat $E(Z(\theta)-Z(\sigma))^2 \le \int_\theta^\sigma q(s)ds < \infty$ voor alle ν en $\theta < \sigma$.

Uit de ongelijkheid van Chebyshev volgt dan dat $Z(\theta) \xrightarrow{P} Z(\sigma)$ als $\theta \to \sigma$ uniform in ν. Met behulp hiervan kan men de uitbreiding van $[0,\sigma)$ naar $[0,\sigma]$ uitvoeren.

Voor de toetsingsgrootheid van Gehan geldt dat

$$E\left(\frac{H_1^2(s)B_1(s)}{\nu}\right) = E\left(\frac{B_2^2(s)B_1(s)}{\nu^3}\right) =$$

$$= \frac{n^2m}{\nu^3}\left[(1-\tilde{G}(s-))^2 + \frac{(1-\tilde{G}(s-))\tilde{G}(s-)}{n}\right](1-\tilde{F}(s-))$$

en

$$E\left(\frac{H_2^2(s)B_2(s)}{\nu}\right) = E\left(\frac{B_1^2(s)B_2(s)}{\nu^3}\right) =$$

$$= \frac{nm^2}{\nu^3}\left[(1-\tilde{F}(s-))^2 + \frac{(1-\tilde{F}(s-))\tilde{F}(s-)}{m}\right](1-\tilde{G}(s-)).$$

Dus

$$E(Z_1(\theta)-Z_1(\sigma))^2 = \sum_{i=1}^{2}\int_\theta^\sigma \nu^{-1}E(H_i^2(s)B_i(s))\lambda(s)ds \le 2\int_\theta^\sigma dF(s) \to 0$$

uniform in ν als $\theta \to \sigma$.

We hebben dus voor de toetsingsgrootheid van Gehan de uitbreiding tot geval D gemaakt.

Voor de toetsingsgrootheid van Cox geldt dat

$$E\left(\frac{H_1^2(s)B_1(s)}{\nu}\right) = E\left(\frac{B_2^2(s)B_1(s)}{(B_1(s)+B_2(s))^2\nu}\right) \le E\left(\frac{B_1(s)}{\nu}\right) = \frac{m}{\nu}\,(1-\tilde{F}(s-))$$

en

$$E\left(\frac{H_2^2(s)B_2(s)}{\nu}\right) \le E\left(\frac{B_2(s)}{\nu}\right) = \frac{n}{\nu}\,(1-\tilde{G}(s-)).$$

Verder geldt dat $\int_0^\sigma (1-\tilde{F}(s-))\lambda(s)ds = \int_0^\sigma (1-J(s-))dF(s) < \infty$, zodat $E(Z_3(\theta)-Z_3(\sigma))^2 \to 0$, uniform in ν, als $\theta \to \sigma$. Hieruit volgt voor de toetsingsgrootheid van Cox de uitbreiding tot de gevallen C en D.

Voor de gemodificeerde toetsingsgrootheid van Efron moet men bewijzen dat $\nu\lambda(s)E(\hat{S}_1^2(s-)\hat{S}_2^2(s-)B_1^{-1}(s)I_{\{B_2(s)>0\}})$ af te schatten is door een integreerbare functie die niet van ν afhangt. Het is ons op dit ogenblik niet bekend of dit mogelijk is.

We gaan nu zwakke konvergentie van de toetsingsgrootheden na bij een rij naburige alternatieven met $\nu = m + n \to \infty$ en $\frac{m}{\nu} \to \rho$ $(0<\rho<1)$.

Stel dat $\lambda_{\nu,1}(s) = \lambda(s)$ (voor alle ν) en $\nu^{\frac{1}{2}}(\lambda_{\nu,2}(s)-\lambda(s)) \to \beta(s)$ als $\nu \to \infty$ voor zekere functie β die aan nader te specificeren eisen moet voldoen.

$$\begin{aligned} Z_\nu(\infty) &= \nu^{\frac{1}{2}} \int_0^\infty J_\nu(s)K_\nu(s)\,\frac{dN_{\nu,2}(s)}{B_{\nu,2}(s)} - \nu^{\frac{1}{2}} \int_0^\infty J_\nu(s)K_\nu(s)\,\frac{dN_{\nu,1}(s)}{B_{\nu,1}(s)} = \\ &= \nu^{\frac{1}{2}} \int_0^\infty \frac{J_\nu(s)K_\nu(s)}{B_{\nu,2}(s)}\,dM_{\nu,2}(s) - \nu^{\frac{1}{2}} \int_0^\infty \frac{J_\nu(s)K_\nu(s)}{B_{\nu,1}(s)}\,dM_{\nu,1}(s) + \\ &\quad + \nu^{\frac{1}{2}} \int_0^\infty J_\nu(s)K_\nu(s)(\lambda_{\nu,2}(s)-\lambda(s))ds \end{aligned}$$

waarbij $J_\nu(s) = I_{\{B_{\nu,1}(s)>0 \text{ én } B_{\nu,2}(s)>0\}}$ en $\{K_\nu(s)\}$ een linkskontinu proces zodanig dat $J_\nu(s)K_\nu(s) \xrightarrow{P} k(s)$ als $\nu \to \infty$.

Merk op dat

$$\nu^{-\frac{1}{2}}H_{\nu,1}(s) = \nu^{\frac{1}{2}}\,\frac{J_\nu(s)K_\nu(s)}{B_{\nu,2}(s)} \quad \text{en} \quad \nu^{-\frac{1}{2}}H_{\nu,2}(s) = \nu^{\frac{1}{2}}\,\frac{J_\nu(s)K_\nu(s)}{B_{\nu,1}(s)}$$

Als

$$\nu^{\frac{1}{2}} \int_0^\infty J_\nu(s)K_\nu(s)(\lambda_{\nu,2}(s)-\lambda(s))ds \xrightarrow{P} \int_0^\infty k(s)\beta(s)ds$$

en bovendien stelling 2.3.1. toepasbaar is, dan zal gelden dat

$$(Z_\nu(\infty) - \int_0^\infty k(s)\beta(s)ds) \xrightarrow{zw} N(0, \int_0^\infty g^2(s)ds)$$

waarbij

$$g^2(s) \overset{\text{def}}{=} g_1^2(s) + g_2^2(s) = \frac{k^2(s)\lambda(s)}{\rho(1-\tilde{F}(s-))} + \frac{k^2(s)\lambda(s)}{(1-\rho)(1-\tilde{G}(s-))} =$$

$$= \rho^{-1}(1-\rho)^{-1} \frac{k^2(s)(1-\tilde{H}(s-))}{(1-\tilde{F}(s-))(1-\tilde{G}(s-))} \overset{\text{def}}{=} k^2(s)\gamma(s)$$

met $\tilde{H}(s-) = \rho\tilde{F}(s-) + (1-\rho)\tilde{G}(s-) = 1 - (1-F(s))(1-L(s-))$.

Onder de nulhypothese hebben wij reeds voor de toetsingsgrootheden van Gehan en Cox en de gemodificeerde toetsingsgrootheid van Efron asymptotische nomaliteit bewezen in een aantal gevallen.

Voor de toetsingsgrootheid van Gehan geldt onder de gemaakte onderstellingen dat

$$J_\nu(s)K_\nu(s) = \frac{B_{\nu,1}(s)B_{\nu,2}(s)}{\nu^2} \xrightarrow{P} \rho(1-\rho)(1-\tilde{F}(s-))(1-\tilde{G}(s-)) = k(s).$$

Voor de gemodificeerde toetsingsgrootheid van Efron geldt onder dezelfde onderstellingen dat

$$J_\nu(s)K_\nu(s) = J_\nu(s)\hat{S}_{\nu,1}(s-)\hat{S}_{\nu,2}(s-) \xrightarrow{P} I_{[0,\sigma]}(s)(1-F(s))^2 = k(s).$$

Tenslotte geldt voor de toetsingsgrootheid van Cox onder de gemaakte onderstellingen

$$J_\nu(s)K_\nu(s) = \frac{1}{\nu} \cdot \frac{B_{\nu,1}(s)B_{\nu,2}(s)}{B_{\nu,1}(s) + B_{\nu,2}(s)} J_\nu(s) \xrightarrow{P} \rho(1-\rho)\frac{(1-\tilde{F}(s-))(1-\tilde{G}(s-))}{1-\tilde{H}(s-)} =$$

$$= k(s) \qquad\qquad \text{op } [0,\sigma].$$

De zwakke konvergentie onder de rij alternatieven is op analoge wijze als onder de nulhypothese te bewijzen.

OPMERKING 3.5.4. In 3.3. van dit hoofdstuk hebben we de varianties van de verschillende toetsingsgrootheden geschat. Men kan nagaan dat de geschatte varianties (met een geschikte normering) in waarschijnlijkheid naar $\int_0^\infty g^2(s)ds$ konvergereren.

Verder kan men nagaan dat de toetsen konsistent zijn.

3.6. ASYMPTOTISCHE RELATIEVE EFFICIENCY EN OPTIMALITEIT

Onder de rij naburige alternatieven geldt in de situatie van de vorige paragraaf dat $Z_\nu(\infty) \xrightarrow{zw} N(\int_0^\infty k(s)\beta(s)ds,\ \int_0^\infty k^2(s)\gamma(s)ds)$ als $\nu \to \infty$ waarbij voor de toetsingsgrootheid van Gehan geldt dat $k(s) = \rho(1-\rho)(1-\tilde{F}(s-))(1-\tilde{G}(s-))$, voor de gemodificeerde toetsingsgrootheid van Efron geldt dat $k(s) = I_{[0,\sigma]}(s)(1-F(s))^2$ en voor de toetsingsgrootheid van Cox geldt dat $k(s) = \rho(1-\rho)\dfrac{(1-\tilde{F}(s-))(1-\tilde{G}(s-))}{(1-\tilde{H}(s-))} I_{[0,\sigma]}(s)$.

$\gamma(s) = \rho^{-1}(1-\rho)^{-1}\dfrac{(1-\tilde{H}(s-))}{(1-\tilde{F}(s-))(1-\tilde{G}(s-))}\lambda(s)$ en $\beta(s) = \lim\limits_{\nu\to\infty} \nu^{\frac{1}{2}}(\lambda_{\nu,2}(s)-\lambda_{\nu,1}(s))$ met $\lambda_{\nu,1}(s) \equiv \lambda(s)$.

De asymptotische relatieve efficiency van de drie toetsen ten opzichte van elkaar is gelijk aan de verhouding tussen

$$\frac{(\int_0^\infty k(s)\beta(s)ds)^2}{\int_0^\infty k^2(s)\gamma(s)ds} = \ell(k)$$

(stelling van Pitman, zie PITMAN [23]).

STELLING 3.6.1. *Stel* $\gamma(s) = 0$ *impliceert* $\beta(s) = 0$. *Definieer* $\dfrac{\beta(s)}{\gamma(s)} = 0$ *als* $\gamma(s) = 0$. *Stel*

$$0 < \int_0^\infty \frac{\beta^2(s)}{\gamma(s)} ds < \infty$$

met β *en* γ *niet-negatieve functies. Definieer* $K = \{k\colon 0 < \int_0^\infty k^2(s)\gamma(s)ds < \infty\}$. *Dan geldt dat*

$$\ell\left(\alpha\,\frac{\beta}{\gamma}\right) = \sup_{k\in K} \ell(k) \qquad \textit{voor alle } \alpha \neq 0.$$

BEWIJS. We merken op dat $\ell(k) = \ell(\alpha k)$ voor alle $\alpha \neq 0$, zodat vanwege de beperking van k dat $\int_0^\infty k^2(s)\gamma(s)ds \in (0,\infty)$ wij kunnen volstaan met het maximaliseren van $\int_0^\infty k(s)\beta(s)ds$ over k onder de voorwaarde dat $\int_0^\infty k^2(s)\gamma(s)ds = c$ met $0 < c < \infty$.

Bekijk nu het probleem:

maximaliseer

$$L(k) = \int_0^\infty k(s)\beta(s)ds - \mu\left(\int_0^\infty k^2(s)\gamma(s)ds - c\right) \text{ over willekeurige } k \in K.$$

Uit de theorie van de Lagrange multiplicatoren volgt dat als k' de oplossing is van dit probleem én $\int_0^\infty k'^2(s)\gamma(s)ds = c$ dan is k' ook de oplossing van het maximaliseren van $\int_0^\infty k(s)\gamma(s)ds$ onder de voorwaarde dat $\int_0^\infty k^2(s)\gamma(s)ds = c$. $L(k) = \int_0^\infty (k(s)\beta(s)-\mu k^2(s)\gamma(s))ds + \mu c$. We kunnen $L(k)$ maximaliseren door de integrand puntsgewijs te maximaliseren. Als $\mu > 0$ dan geeft dit $k(s) = \frac{\beta(s)}{2\mu\gamma(s)}$ voor alle s.

Dus als $\mu > 0$, $c > 0$ zo gekozen zijn dat $4\mu^2 c = \int_0^\infty \frac{\beta^2(s)}{\gamma(s)} ds$ dan is $L(k)$ gemaximaliseerd over willekeurige $k \in K$, door $k(s) = \frac{\beta(s)}{2\mu\gamma(s)}$. Deze oplossing voldoet tevens aan $\int_0^\infty k^2(s)\gamma(s)ds = c$. □

OPMERKING 3.6.2. De toets van Cox (ook wel log rank test genoemd) heeft een belangrijke eigenschap.

Stel $\lambda_{\nu,1}(s) = \lambda(s)$, $\beta(s) = \beta\lambda(s)$ en $\lambda_{\nu,2}(s) = \lambda(s)(1+\nu^{-\frac{1}{2}}\beta)$ (dit is een zogenaamde rij van Lehmann-alternatieven die naar de nulhypothese konvergeren.

Uit de vorige stelling weten we dat de optimale k evenredig is met

$$\frac{\beta(s)}{\gamma(s)} = \beta\frac{(1-\tilde{F}(s-))(1-\tilde{G}(s-))}{(1-\tilde{H}(s-))} \text{ op } [0,\sigma].$$

Hieruit volgt dus dat de toets van Cox optimaal is bij Lehmann-alternatieven in de klasse van toetsen die in ons model passen.

VOORBEELD 3.5.3. We zullen een efficiency berekening maken voor de logistieke verdeling. Dus

$$F_0(x) = (1+e^{-x})^{-1}$$
$$f_0(x) = e^{-x}(1+e^{-x})^{-2}$$

en

$$\lambda_0(x) = (1+e^{-x})^{-1}$$

De toetsen van Gehan en Efron zijn generalisaties van de toets van Wilcoxon. We verwachten daarom een hoge efficiency in een locatie familie van logistieke verdelingen (die bij 0 links afgeknot zijn). Zij $F_\theta(x) = F(x-\theta)$, $f_\theta(x) = F'_\theta(x)$ en $\lambda_\theta(x) = \frac{f_\theta(x)}{1-F_\theta(x)}$. Er geldt dat $\frac{d}{d\theta}\lambda_\theta(x) = -f_\theta(x)$.
Beschouw nu een rij steekproeven (die binnen het model passen) met onderliggende verdelingen de logistieke verdeling met parameters $\theta_{\nu,1} \equiv \theta$ en $\theta_{\nu,2}$, waarbij

$$\nu^{\frac{1}{2}}(\theta-\theta_{\nu,2}) \to c > 0$$

en

$$\nu^{\frac{1}{2}}(\lambda_{\theta_{\nu,2}}(x)-\lambda_\theta(x)) \to cf_\theta(x) \qquad \text{als } \nu \to \infty.$$

De grootste efficiency kunnen wij verwachten als $k(s)$ evenredig is met $\frac{\beta(s)}{\gamma(s)}$ op $[0,\sigma]$ met $\beta(s) = f_\theta(s)$ en $\gamma(s) = \frac{(1-\tilde{H}(s-))\lambda_\theta(s)}{(1-\tilde{F}(s-))(1-\tilde{G}(s-))}$.

De optimale k(s) zou dus evenredig moeten zijn met $\frac{(1-F(s))(1-\tilde{F}(s-))(1-\tilde{G}(s-))}{(1-\tilde{H}(s-))}$ op $[0,\sigma]$.

Bij type I censurering (zie 1.1.) hebben wij dat k(s) evenredig is met $(1-F(s))^2$ omdat $\tilde{F}(s) = \tilde{G}(s) = \tilde{H}(s) = F(s)$ voor $s \in [0,\sigma]$.

De toetsingsgrootheid van Gehan (die in dit geval gelijk is aan de gemodificeerde toetsingsgrootheid van Efron) is in deze situatie optimaal.

Bij meer algemene censurering lijkt een andere toetsingsgrootheid aangewezen, bijvoorbeeld

$$Z(\infty) = \int_0^\infty \frac{K(s)}{B_2(s)}\, dN_2(s) - \int_0^\infty \frac{K(s)}{B_1(s)}\, dN_1(s)$$

waarbij

$$K(s) = \frac{\hat{S}(s-)B_1(s)B_2(s)}{B_1(s) + B_2(s)}$$

met $\hat{S}(s-)$ de produkt-limiet schatter gebaseerd op de gekombineerde steekproef. ($\hat{S}(s-) = \frac{m}{\nu}\hat{S}_1(s-) + \frac{n}{\nu}\hat{S}_2(s-)$ waarin $\hat{S}_1(s-)$ de produkt-limiet schatter is uit de eerste steekproef en $\hat{S}_2(s-)$ de produkt-limiet schatter is uit de tweede steekproef.)

LITERATUUR

[1] AALEN, O.O. (1976), *Statistical inference for a family of counting processes*, Ph.D. thesis, University of Copenhagen.

[2] AALEN, O.O. (1976), *On nonparametric tests for comparison of two counting processes*, Working paper no. 6, Laboratory of Actuarial Mathematics, University of Copenhagen.

[3] AALEN, O.O. (1977), *Weak convergence of stochastic integrals related to counting processes*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 38, 261-277.

[4] BILLINGSLEY, P. (1968), *Convergence of probability measures*, Wiley, New York.

[5] BOEL, R., P. VARAIYA & E. WONG (1975), *Martingales on jump processes I: Representation results*, SIAM J. Control 13, 999-1021.

[6] BRESLOW, N. & J. CROWLEY (1974), *A large sample study of the life table and product limit estimates under random censorship*, Ann. Statist. 2, 437-453.

[7] COX, D.R. (1972), *Regression models and life-tables*, J. Roy. Statist. Soc. B 34, 187-220.

[8] DELLACHERIE, C. (1972), *Capacités et processus stochastiques*, Ergebnisse der Mathematik usw. 67, Springer-Verlag, Berlin.

[9] DIEUDONNÉ, J. (1970), *Treatise on analysis*, Academic Press, New York.

[10] DOLÉANS, C. (1967), *Processus croissants naturels et processus croissant très-bien-mesurables*, C.R. Acad. Sc. Paris 264, 874-876.

[11] DOLÉANS-DADE, C. & P.A. MEYER (1970), *Intégrales stochastiques par rapport aux martingales locales*, Séminaire de Probabilités, Lecture Notes in Mathematics 124, 77-107, Springer-Verlag, Berlin.

[12] DOLIVO, F.B. (1974), *Counting processes and integrated conditional rates: A martingale approach with application to detection*, Technical Report, College of Engineering, University of Michigan.

[13] EFRON, B. (1967), *The two sample problem with censored data*, Proc. Fifth Berkeley Symp. Math. Stat. Prob. 4, 831-853.

[14] GEHAN, E.A. (1965), *A generalized Wilcoxon test for comparing arbitrarily singly-censored samples*, Biometrika 52, 203-223.

[15] HOEFFDING, W. (1956), *On the distribution of the number of successes in independent trials*, Ann. Math. Statist. 27, 713-721.

[16] KAPLAN, E.L. & P. MEIER (1958), *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. 53, 457-481.

[17] LOÈVE, M. (1963), *Probability theory*, Van Nostrand, Princeton.

[18] MANTEL, N. (1967), *Ranking procedures for arbitrarily restricted observation*, Biometrics 23, 65-78.

[19] MEYER, P.A. (1966), *Probabilités et potentiel*, Hermann, Paris.

[20] MEYER, P.A. (1976), *Un cours sur les integrales stochastiques*, Séminaire de Probabilités, Lecture Notes in Mathematics 511, 245-400, Springer-Verlag, Berlin.

[21] NELSON, W. (1969), *Hazard plotting for incomplete failure data*, J. Quality Techn. 1, 27-52.

[22] PETO, R. & J. PETO (1972), *Asymptotically efficient rank invariant test procedures*, J. Roy. Statist. Soc. A 135, 185-206.

[23] PITMAN, E.J.G. (1949), *Lecture notes on nonparametric statistical inference*, unpublished.

[24] PURI, M.L. & P.K. SEN (1971), *Nonparametric methods in multivariate analysis*, Wiley, New York.

[25] THOMAS, D.R. (1969), *Conditionally locally most powerful rank tests for the two-sample problem with arbitrarily censored data*, Technical Report 7, Department of Statistics, Oregon State University.

[26] TUCKER, H.G. (1967), *A graduate course in probability*, Academic Press, New York.

[27] VERVAAT, W. (1972), *Success epochs in Bernoulli trials (with applications in number theory)*, M.C. Tracts 42, Amsterdam.

[28] ZAANEN, A.C. (1967), *Integration*, North-Holland Publishing Company, Amsterdam.

[29] ZUIJLEN, M.C.A. VAN (1977), *Empirical distributions and rankstatistics*, M.C. Tracts 79, Amsterdam.