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).

MC SYLLABUS 22

R. HELMERS
J. OOSTERHOFF
F.H. RUYMGAART
M.C.A. VAN ZUYLEN

ASYMPTOTISCHE METHODEN IN DE TOETSINGSTHEORIE

TOEPASSINGEN VAN NABURIGHEID

MATHEMATISCH CENTRUM AMSTERDAM 1976

AMS(MOS) subject classification scheme (1970): 60F05, 62E20, 62G10, 62G20

ISBN 90 6196 104 1

INHOUD

Voorwoord	i
Inleiding	1 - 3
Hoofdstuk 1: Naburige rijen kansmaten	4 - 31
Hoofdstuk 2: Asymptotische toetsingstheorie	32 - 56
Hoofdstuk 3: Asymptotische optimaliteit van rangtoetsen	58 - 112
Vraagstukken	113 - 155
Literatuur	156 - 157
Symbolen	158 - 160
Index	161 - 164

VOORWOORD

Naar aanleiding van een in augustus 1973 door het Mathematisch Centrum georganiseerde cursus over asymptotische methoden in de statistiek is, als leidraad voor de deelnemers, deze syllabus geschreven. Bij de samenstelling is getracht een op zichzelf staande inleiding te geven tot de asymptotische toetsingstheorie in het algemeen en tot de voor de praktijk belangrijke asymptotische theorie van rangtoetsen in het bijzonder.

Het karakter van deze syllabus is in belangrijke mate bepaald door de aard van de cursus waarvoor hij werd geschreven. In de tekst worden slechts algemeen bekende resultaten gepresenteerd zonder dat naar volledigheid is gestreefd. Het was veeleer de bedoeling van deze cursus om in kort bestek een uiteenzetting te geven van de grondbeginselen van de theorie die bij practische toepassingen een grote rol spelen. Met het oog op verdere studie is aan ieder hoofdstuk een paragraaf met bronvermeldingen toegevoegd. Ook is een serie uitgewerkte vraagstukken opgenomen.

Hoewel de inhoud van de syllabus een afgerond geheel vormt, wordt van de lezer vertrouwdheid verwacht met toepassing van wiskundige technieken. Enige kennis van maattheorie en waarschijnlijkheidsrekening is onontbeerlijk. Gezien het theoretisch karakter van deze cursus is enig inzicht in de statistiek en in het bijzonder in de toetsingstheorie wenselijk.

Amsterdam, 1976

R. Helmers,
J. Oosterhoff,
F.H. Ruymgaart,
M.C.A. van Zuylen.

INLEIDING

In de praktijk is de statisticus herhaaldelijk genoodzaakt een beroep te doen op asymptotische methoden, aangezien exacte procedures onuitvoerbaar of zelfs onbekend blijken te zijn. In de verdelingsvrije statistiek vindt men bij uitstek voorbeelden van deze situatie. Wij zullen ons hier tot de asymptotische toetsingstheorie beperken. Deze behelst onder meer de constructie van asymptotisch meest onderscheidende toetsingsprocedures, de bepaling van de asymptotische kansverdeling van de toetsingsgrootheden en het vergelijken van twee procedures door middel van hun asymptotische relatieve doeltreffendheid. Het blijkt nu dat vele verdelingsvrije toetsen die ingang vonden op grond van eenvoud van interpretatie en berekeningswijze, bovendien asymptotisch goede eigenschappen bezitten.

Tot de specifieke problemen van de asymptotische toetsingstheorie behoort het definieren van een zinvol optimaliteitscriterium. Omdat bij vaste nulhypothese, vast alternatief en onbegrensd toenemende steekproefomvang een rij toetsen, ook al is deze geenszins optimaal, dikwijls wel consistent is, levert de limiet van het onderscheidingsvermogen in deze situatie geen zinvolle maatstaf op. Laten we daarom de eigenschappen van de rij toetsen onderzoeken voor alternatieven "in de buurt van" de nulhypothese. Daartoe kiezen we voor iedere steekproefomvang één kansmaat uit de nulhypothese en één uit het alternatief en wel zó dat deze kansmaten, bij toenemende steekproefomvang, in zekere zin tot elkaar naderen. In het algemeen zal de limiet van het onderscheidingsvermogen van de toetsen in de rij nu een getal kleiner dan 1 zijn. Deze limiet zal doorgaans voor verschillende rijen toetsen verschillend zijn, zodat in de waarde van de limiet de kwaliteit van de rij toetsen tot uitdrukking komt.

Laten wij thans tot een iets nauwkeuriger formulering overgaan. Stel dat voor ieder $N = 1,2,\ldots$ een steekproef van de omvang N gegeven is, waarvan de steekproefelementen alle een meetbare ruimte (X,A) als uitkomstenruimte hebben. De uitkomstenruimte van de N-de steekproef is dus (X^N,A^N). Zij nu P_N een kansverdeling op (X^N,A^N) uit de nulhypothese, Q_N een kansverdeling op (X^N,A^N) uit het alternatief en μ_N een σ-eindige maat, die P_N en Q_N domineert (b.v. $\mu_N := P_N + Q_N$). Noteer de dichtheid van $P_N(Q_N)$

t.o.v. μ_N met $p_N(q_N)$. De voorwaarde dat de kansmaten op een geschikte manier tot elkaar naderen wordt formeel weergegeven door de eis dat de rijen $\{P_N\}$ en $\{Q_N\}$ naburig zijn t.o.v. elkaar, d.w.z. door de eis dat (voor $A_N \in \mathcal{A}_N$) de equivalentie

$$(0.1) \qquad [\lim_{N\to\infty} P_N(A_N) = 0] \iff [\lim_{N\to\infty} Q_N(A_N) = 0]$$

geldt.

In hoofdstuk 1 geven we een algemene inleiding tot het begrip *naburigheid*, dat in de asymptotische statistiek een wezenlijke rol speelt. In het bijzonder geven we criteria voor de geldigheid van (0.1) in termen van de logarithme log L_N van het aannemelijkheidsquotient L_N, gedefinieerd als

$$(0.2) \qquad L_N(x) = \begin{cases} q_N(x) \,/\, p_N(x) & \text{als } p_N(x) > 0, \\ \infty & \text{als } p_N(x) = 0. \end{cases}$$

De grondbeginselen van de asymptotische toetsingstheorie zetten we in hoofdstuk 2 uiteen. We beschouwen daar de situatie dat voor iedere N de enkelvoudige nulhypothese P_N tegen het enkelvoudige alternatief Q_N getoetst wordt. Op grond van het lemma van Neyman-Pearson bestaat er voor iedere N een meest onderscheidende toets ϕ_N^*. Aangezien dit lemma in het vervolg een wezenlijke rol speelt zullen wij het volledigheidshalve hier vermelden.

LEMMA van NEYMAN-PEARSON: *Zij* α *een vast getal zodat* $0 < \alpha < 1$. *Voor het toetsen van de enkelvoudige nulhypothese* P_N *tegen het enkelvoudige alternatief* Q_N, *met onbetrouwbaarheidsdrempel* α, *is de toets*

$$(0.3) \qquad \phi_N^*(x) := \begin{cases} 1 & \text{als } L_N(x) > k, \\ \gamma & \text{als } L_N(x) = k, \\ 0 & \text{als } L_N(x) < k, \end{cases}$$

waarin men $k(0 \le k \le \infty)$ en $\gamma(0 \le \gamma \le 1)$ *zó moet (en kan) bepalen dat*

(0.4) $$P_N(\{L_N>k\}) + \gamma P_N(\{L_N=k\}) = \alpha,$$

meest onderscheidend.

In hoofdstuk 2 voeren we het begrip *asymptotisch meest onderscheidende* rij toetsen (met asymptotische onbetrouwbaarheidsdrempel α) in en bewijzen dat de rij $\{\phi_N^*\}$ van bovenstaande toetsen asymptotisch meest onderscheidend is, hetgeen intuïtief voor de hand ligt. Het belangrijkste resultaat van dit hoofdstuk is echter dat men, onder zekere voorwaarden op de rijen $\{P_N\}$ en $\{Q_N\}$ die naburigheid impliceren, eenvoudiger toetsen kan construeren die eveneens asymptotisch meest onderscheidend zijn. Bovendien kunnen we de asymptotische verdeling van de toetsingsgrootheden, waarop deze toetsen gebaseerd zijn, zowel onder $\{P_N\}$ als onder $\{Q_N\}$ bepalen. Ook het asymptotisch onderscheidingsvermogen kunnen we derhalve berekenen. Het bewijs berust op een vergelijking met de rij toetsen $\{\phi_N^*\}$.

Hoofdstuk 3 tenslotte is gewijd aan de constructie van *asymptotisch uniform meest onderscheidende* rangtoetsen voor samengestelde nulhypothesen tegen samengestelde alternatieven (met asymptotische onbetrouwbaarheidsdrempel α). Ook nu wordt weer de asymptotische verdeling van de toetsingsgrootheden, waarop deze rangtoetsen zijn gebaseerd, onder nulhypothese en naburig alternatief berekend. Het bewijs berust op een vergelijking met rijen asymptotisch meest onderscheidende (maar niet verdelingsvrije) toetsen die volgens de methode van hoofdstuk 2 geconstrueerd zijn. Voor een drietal belangrijke problemen uit de verdelingsvrije toetsingstheorie, die veelvuldig in de praktijk voorkomen, bepalen we aldus rangtoetsen welke tegen bepaalde samengestelde alternatieven asymptotisch uniform meest onderscheidend zijn, nl.:

(a) het regressie probleem (meer in het bijzonder het twee- en meer-steekproeven probleem),
(b) het symmetrie probleem,
(c) het onafhankelijkheidsprobleem.

Bekende toetsen zoals de twee-steekproeven toets van Wilcoxon, de symmetrie toets van Wilcoxon en de rangcorrelatie toets van Spearman blijken tegen bepaalde alternatieven asymptotisch uniform meest onderscheidende toetsen te zijn.

Hoofdstuk 1 NABURIGE RIJEN KANSMATEN

In dit hoofdstuk bespreken we de definitie en een aantal karakteriseringen van naburige (Engels: contiguous) rijen kansmaten. In de volgende twee hoofdstukken zullen we enkele toepassingen laten zien.

De theorie werd voor het eerst opgezet door LECAM ([16],[17]). Min of meer uitvoerige beschouwingen treft men o.a. ook aan in HÁJEK [10], HÁJEK & ŠIDÁK [12], ROUSSAS [22] en WITTING & NÖLLE [24].

In het vervolg zal convergentie van kansverdelingen van stochastische grootheden een grote rol spelen. Daar deze stochastische grootheden dikwijls met positieve kans de waarden $\pm\infty$ aannemen, zijn deze kansverdelingen vaak defectieve kansmaten op $\mathbb{R}$. Het meest voor de hand liggende convergentiebegrip in dergelijke situaties is vage convergentie (zie b.v. CHUNG [7]). We geven daarom volledigheidshalve eerst een korte uiteenzetting over vage convergentie en meer in het bijzonder over zwakke convergentie.

We beginnen met het vastleggen van enige notatie en begrippen. De Borel-σ-algebra van deelverzamelingen van $\mathbb{R}^k$ geven we aan met $\mathcal{B}^k$ ($k \geq 1$), waarbij we $\mathcal{B}^1$ als $\mathcal{B}$ schrijven. Het complement van een verzameling B duiden we aan met B^c, de afsluiting van B met $\bar{B}$ en de rand van B met ∂B. De indicatorfunctie χ_B van een verzameling B zullen we vaak noteren met $\chi(B)$. Voorts is $C_B(S)$ de klasse van alle reëelwaardige begrensde continue functies op S en is $C_C(S)$ de klasse van alle reëelwaardige continue functies op S met compacte drager (S is een willekeurige topologische ruimte).

Een maat P op $(\mathbb{R}^k,\mathcal{B}^k)$ met $P(\mathbb{R}^k) \leq 1$ noemen we een *defectieve kansmaat*. De defectieve kansmaten omvatten dus de kansmaten. Met een defectieve kansmaat P op $(\mathbb{R}^k,\mathcal{B}^k)$ correspondeert 1-1 duidig een defectieve verdelingsfunctie F gedefinieerd door

$$F(x) := P((-\infty,x]) \qquad \text{voor alle } x \in \mathbb{R}^k.$$

In het algemeen is $\lim_{x\to\infty} F(x) \leq 1$, maar als P een kansmaat is, dan is $\lim_{x\to\infty} F(x) = 1$ en is F dus een verdelingsfunctie.

Stochastische grootheden (met waarden in $\mathbb{R}$ of $\bar{\mathbb{R}}$) en stochastische vectoren worden steeds aangegeven met hoofdletters; zo zullen L, Λ, R, S, T, U, V, W, X, Y en Z, met of zonder indices, steeds stochastische grootheden (vectoren) voorstellen. We staan toe dat stochastische grootheden (vectoren)

met positieve kans oneindige waarden aannemen. Met de defectieve kansverdeling van een stochastische k-vector ($k \geq 1$) X bedoelen we steeds de door X geïnduceerde defectieve kansmaat op $(\mathbb{R}^k, \mathcal{B}^k)$. We duiden deze dikwijls aan met $L(X)$.

De normale verdeling met verwachting μ en variantie σ^2 geven we aan met $N(\mu;\sigma^2)$; de 2-dimensionale normale verdeling met verwachtingen μ en ν, varianties σ^2 en τ^2, en correlatiecoëfficiënt ρ schrijven we als $N(\mu,\nu;\sigma^2,\tau^2,\rho)$. De standaard normale verdelingsfunctie geven we aan met Φ en $\xi_{1-\alpha}$ is gedefinieerd door $\Phi(\xi_{1-\alpha}) = 1-\alpha$.

Integreren we een reëelwaardige meetbare functie h van een stochastische k-vector X naar de (defectieve) kansverdeling $L(X)$, dan verstaan we onder de integraal $\int h(x)dL(x)$ steeds de integraal over het integratiegebied $\mathbb{R}^k$.

We introduceren thans eerst het begrip vage convergentie.

DEFINITIE 1.1: Zijn P_N (N=1,2,...) en P defectieve kansmaten op $(\mathbb{R}^k, \mathcal{B}^k)$, dan zeggen we dat P_N *vaag convergeert* naar P, notatie $P_N \rightarrow_v P$, als

$$\lim_{N\to\infty} \int f dP_N = \int f dP \qquad \text{voor alle } f \in C_C(\mathbb{R}^k). \tag{1.1}$$

Merk op dat uit $P_N \rightarrow_v P$ *niet* volgt $P_N(\mathbb{R}^k) \rightarrow P(\mathbb{R}^k)$. De betekenis van vage convergentie komt duidelijk tot uitdrukking in het volgende karakteriseringslemma.

LEMMA 1.1: Zijn P_N (N=1,2,...) en P defectieve kansmaten op $(\mathbb{R}^k, \mathcal{B}^k)$, dan zijn de volgende uitspraken equivalent:

(a) $P_N \rightarrow_v P$,

(b) $\limsup_{N\to\infty} P_N(K) \leq P(K)$ voor alle begrensde gesloten $K \subset \mathbb{R}^k$

én $\liminf_{N\to\infty} P_N(G) \geq P(G)$ voor alle begrensde open $G \subset \mathbb{R}^k$,

(c) $\lim_{N\to\infty} P_N(B) = P(B)$ voor alle begrensde $B \in \mathcal{B}^k$ met $P(\partial B) = 0$.

BEWIJS: Zie HEYER [13] p. 62. □

In het vervolg zullen we steeds te maken krijgen met de situatie waarin de limiet P van een rij defectieve kansmaten $\{P_N\}$ zelf een kansmaat is. In dit geval geldt het volgende resultaat.

LEMMA 1.2: Is $\{P_N\}$ een rij defectieve kansmaten en P een kansmaat op $(\mathbb{R}^k, \mathcal{B}^k)$, dan impliceert $P_N \rightarrow_v P$ dat

$$\lim_{N\to\infty} \int f dP_N = \int f dP \qquad \text{voor alle } f \in C_B(\mathbb{R}^k). \tag{1.2}$$

BEWIJS:
Zij $\varepsilon > 0$. Dan bestaat er een compacte $K \subset \mathbb{R}^k$ zodat $P(K) > 1-\varepsilon$. Op grond van Urysohn's lemma (zie b.v. RUDIN [23] p. 39) bestaat er een functie $h \in C_C(\mathbb{R}^k)$ met $0 \leq h \leq 1$ en $h(x) = 1$ voor alle $x \in K$. Dan is

$$P_N(\mathbb{R}^k) \geq \int h dP_N \to \int h dP \geq P(K) > 1-\varepsilon$$

wegens (1.1), zodat $\lim_{N\to\infty} P_N(\mathbb{R}^k) = 1$.

Zij nu $f \in C_B(\mathbb{R}^k)$ en onderstel dat $0 \leq f \leq M$ voor zekere constante M. Definieer $g := h.f$, dan is $g \in C_C(\mathbb{R}^k)$, $0 \leq g \leq f$ en $g = f$ op K. Dit impliceert

$$\int f dP = \int_K g dP + \int_{K^c} f dP \leq \int g dP + MP(K^c) < \int g dP + M\varepsilon.$$

Uit deze ongelijkheid en (1.1) volgt voor alle $N \geq N'_\varepsilon$

$$\int f dP_N \geq \int g dP_N > \int g dP - \varepsilon > \int f dP - (M+1)\varepsilon.$$

Zij thans $f^* := M-f$ en ga op dezelfde wijze te werk met f^*; dit geeft

$$\int f^* dP_N > \int f^* dP - (M+1)\varepsilon \qquad \text{voor alle } N \geq N''_\varepsilon,$$

of wel

$$\int f dP_N < \int f dP + M\{P_N(\mathbb{R}^k)-1\} + (M+1)\varepsilon \leq \int f dP + (M+1)\varepsilon$$

voor alle $N \geq N''_\varepsilon$. Maar dan is

$$\left|\int f dP_N - \int f dP\right| < (M+1)\varepsilon \qquad \text{voor alle } N \geq \max\{N'_\varepsilon, N''_\varepsilon\},$$

waarmee bewezen is dat $\lim_{N\to\infty} \int f dP_N = \int f dP$ voor alle $f \in C_B(\mathbb{R}^k)$ met $f \geq 0$.

Als niet voldaan is aan $f \geq 0$, maar $f \geq -M$ $(M>0)$, beschouw dan $\tilde{f} := f+M$. Dan volgt uit het voorgaande dat

$$\lim_{N\to\infty} \int f dP_N = \lim_{N\to\infty} \int (\tilde{f}-M) dP_N =$$

$$= \int \tilde{f} dP - M \lim_{N\to\infty} P_N(\mathbb{R}^k) = \int f dP + M - M = \int f dP. \qquad \square$$

OPMERKING 1.1: In de literatuur ziet men voor defectieve kansmaten $\{P_N\}$ en P meestal bewezen dat $P_N \to_v P$ en $P_N(\mathbb{R}^k) \to P(\mathbb{R}^k)$ tezamen de relatie (1.2) impliceren. Zoals uit het bewijs van lemma 1.2 blijkt, is de conditie $P_N(\mathbb{R}^k) \to P(\mathbb{R}^k)$ automatisch vervuld als P een kansmaat is.

In verband met het veelvuldig optreden van vage convergentie naar een kansmaat geven we hieraan een aparte naam: zwakke convergentie.

DEFINITIE 1.2: Is $\{P_N\}$ een rij defectieve kansmaten en P een kansmaat op $(\mathbb{R}^k, \mathcal{B}^k)$, dan zeggen we dat P_N *zwak convergeert* naar P, notatie $P_N \to_z P$, als (1.2) vervuld is.

Bij zwakke convergentie krijgt lemma 1.1 een iets andere gedaante.

LEMMA 1.3. Is $\{P_N\}$ een rij defectieve kansmaten en P een kansmaat op $(\mathbb{R}^k, \mathcal{B}^k)$, dan zijn de volgende uitspraken equivalent:

(i) $P_N \to_z P$,

(ii) $\limsup_{N\to\infty} P_N(F) \leq P(F)$ voor alle gesloten $F \subset \mathbb{R}^k$

én $\lim_{N\to\infty} P_N(\mathbb{R}^k) = 1$,

(iii) $\liminf_{N\to\infty} P_N(G) \geq P(G)$ voor alle open $G \subset \mathbb{R}^k$,

(iv) $\lim_{N\to\infty} P_N(B) = P(B)$ voor alle $B \in \mathcal{B}^k$ met $P(\partial B) = 0$.

BEWIJS:
Voor kansmaten $\{P_N\}$ en P wordt dit lemma o.a. bewezen in BILLINGSLEY [6] p. 12-14 en in FABIUS & VAN ZWET [8] p. 105-108. Deze bewijzen gaan nagenoeg ongewijzigd door in het geval $\{P_N\}$ een rij defectieve kansmaten is, mits $P_N(\mathbb{R}^k) \to 1$. In (i), (iii) en (iv) volgt deze relatie direct uit de daar genoemde eigenschappen en het feit dat P een kansmaat is, maar in (ii) moet het afzonderlijk geëist worden. □

OPMERKING 1.2: De tot dusverre besproken theorie blijft juist als men $(\mathbb{R}^k,\mathcal{B}^k)$ vervangt door een willekeurige locaal-compacte separabele metrische ruimte met de σ-algebra van Borelverzamelingen. Zelfs is generalisatie tot nog algemenere ruimten mogelijk; zie b.v. BILLINGSLEY [6], HEYER [13] of BAUER [2]. We zullen hier echter geen gebruik van maken.

Het volgende lemma geeft nog iets meer informatie over zwakke convergentie.

LEMMA 1.4: Zij $\{P_N\}$ een rij defectieve kansmaten en P een kansmaat op $(\mathbb{R}^k,\mathcal{B}^k)$, zij h een reëelwaardige meetbare functie op $\mathbb{R}^k$ en zij D_h de verzameling van alle discontinuïteitspunten van h. Laat voorts P_N^h (N=1,2,...) en P^h de door h geïnduceerde (defectieve) kansmaten op $(\mathbb{R},\mathcal{B})$ zijn. Dan volgt uit $P_N \to_z P$ dat

(1.3) $P_N^h \to_z P^h$ mits $P(D_h) = 0$, en

(1.4) $\lim_{N\to\infty} \int h dP_N = \int h dP$ mits h begrensd is en $P(D_h) = 0$.

BEWIJS:
Het in BILLINGSLEY [6] p. 31 gegeven bewijs voor kansmaten $\{P_N\}$ en P gaat ongewijzigd door. □

Kiezen we voor de functie h de projectie van $\mathbb{R}^k$ (k>1) op één van de coördinaten, dan blijkt uit lemma 1.4 dat bij zwakke convergentie van P_N naar P in $(\mathbb{R}^k,\mathcal{B}^k)$ de marginale maten zwak meeconvergeren, aangezien een projectie een continue afbeelding is.

DEFINITIE 1.3: Een rij defectieve kansmaten $\{P_N\}$ op $(\mathbb{R}^k,\mathcal{B}^k)$ heet *relatief compact* als elke deelrij een subdeelrij bevat die zwak convergeert (naar een kansmaat die niet tot $\{P_N\}$ behoeft te behoren).

DEFINITIE 1.4: Een rij defectieve kansmaten $\{P_N\}$ op $(\mathbb{R}^k,\mathcal{B}^k)$ heet *uniform beperkt* (Engels: *tight*) als bij elke $\varepsilon > 0$ een compacte $K_\varepsilon \subset \mathbb{R}^k$ en een index N_ε bestaat zó dat $P_N(K_\varepsilon) > 1-\varepsilon$ voor alle $N \geq N_\varepsilon$.

Definitie 1.4 wijkt enigszins af van de gebruikelijke definitie van uniforme beperktheid, waarbij gewoonlijk geëist wordt dat $P_N(K_\varepsilon^c) < \varepsilon$ voor alle N. Deze eis is voor onze doeleinden echter te zwak, omdat hieruit niet volgt dat $P_N(\mathbb{R}^k) \to 1$ voor $N \to \infty$. Deze laatste voorwaarde speelt in het vervolg een essentiële rol (vergelijk ook definitie 1.3). Als alle P_N kansmaten zijn, dan stemt de definitie echter overeen met die gegeven in b.v. BILLINGSLEY [6] en FABIUS & VAN ZWET [8].

Het volgende lemma legt een verband tussen relatieve compactheid en uniforme beperktheid.

LEMMA 1.5: Zij $\{P_N\}$ een rij defectieve kansmaten op $(\mathbb{R}^k,\mathcal{B}^k)$. Dan is $\{P_N\}$ relatief compact d.e.s.d. als $\{P_N\}$ uniform beperkt is.

BEWIJS:
Het in BILLINGSLEY [6] p. 37-40 en FABIUS & VAN ZWET [8] p. 113 gegeven bewijs blijft met kleine wijzigingen doorgaan. □

Zwakke convergentie van defectieve kansmaten op $(\mathbb{R}^k,\mathcal{B}^k)$ kan ook uitgedrukt worden als zwakke convergentie van de corresponderende defectieve verdelingsfuncties.

DEFINITIE 1.5: Zij $\{F_N\}$ een rij defectieve verdelingsfuncties en F een verdelingsfunctie op $\mathbb{R}^k$. Dan zeggen we dat F_N *zwak convergeert* naar F, notatie $F_N \longrightarrow_z F$, als

$$\lim_{N\to\infty} F_N(x) = F(x) \quad \text{voor alle continuïteitspunten } x \text{ van } F.$$

LEMMA 1.6: Zij $\{P_N\}$ een rij defectieve kansmaten op $(\mathbb{R}^k,\mathcal{B}^k)$ met defectieve verdelingsfuncties $\{F_N\}$. Zij P een kansmaat op $(\mathbb{R}^k,\mathcal{B}^k)$ met verdelingsfunctie F. Dan geldt: $P_N \longrightarrow_z P$ d.e.s.d. als $F_N \longrightarrow_z F$.

BEWIJS:
Dit bewijs verloopt geheel analoog aan het bewijs gegeven in BILLINGSLEY [6] p. 17-18 en FABIUS & VAN ZWET [8] p. 105-108. □

Met behulp van zwakke convergentie van defectieve kansmaten kunnen we nu ook convergentie in verdeling van stochastische grootheden (vectoren) definiëren.

DEFINITIE 1.6: Zij $\{X_N\}$ een rij stochastische k-vectoren ($k \geq 1$), niet noodzakelijk op eenzelfde kansruimte, met defectieve kansverdelingen $\{L(X_N)\}$. Dan zeggen we dat X_N *in verdeling convergeert* naar een kansmaat P op $(\mathbb{R}^k,\mathcal{B}^k)$, notatie $X_N \longrightarrow_D P$, als $L(X_N) \longrightarrow_z P$. Is X een stochastische k-vector met P als kansverdeling, dan zeggen we ook wel dat X_N in verdeling convergeert naar X, notatie $X_N \longrightarrow_D X$.

Op grond van de eerste uitspraak van lemma 1.4 impliceert convergentie in verdeling van stochastische k-vectoren X_N naar X ($k>1$) dat elke lineaire combinatie van de coördinaten van X_N in verdeling convergeert naar de overeenkomstige lineaire combinatie van de coördinaten van X. Deze eigenschap heeft ook een omkering, die vaak handig toepasbaar is en bekend staat als het "Cramér-Wold device".

LEMMA 1.7: Zijn $\{X_N\}$ en X_0 stochastische k-vectoren ($k>1$) met coördinaten $X_{N,i}$ en $X_{0,i}$ ($i=1,2,\ldots,k$), dan zijn de beide volgende uitspraken equiva-

lent:

(i) $X_N \rightarrow_D X_0$,

(ii) voor elke constante vector $c \in \mathbb{R}^k$ geldt dat

$$\sum_{i=1}^k c_i X_{N,i} \rightarrow_D \sum_{i=1}^k c_i X_{0,i}.$$

BEWIJS:
Het bewijs in BILLINGSLEY [6] p. 48-49 gaat ongewijzigd door; zie ook WITTING & NÖLLE [24] p. 54. □

Van groot belang is ook convergentie in waarschijnlijkheid van een rij stochastische grootheden (vectoren) naar een constante. We geven een iets algemenere definitie dan gebruikelijk.

DEFINITIE 1.7: Zij X_N een stochastische k-vector ($k \geq 1$) op een kansruimte (Ω_N, A_N, P_N), $N = 1,2,\ldots$, en zij $a \in \mathbb{R}^k$ een constante. Dan zeggen we dat X_N *in waarschijnlijkheid convergeert* naar a, notatie $X_N \rightarrow_{P_N} a$, als

$$\lim_{N\to\infty} P_N(\|X_N - a\| > \varepsilon) = 0 \qquad \text{voor alle } \varepsilon > 0, \tag{1.5}$$

waarin $\|.\|$ de euclidische afstand in $\mathbb{R}^k$ voorstelt.

Ook bij deze definitie blijft het algemeen bekende resultaat doorgaan dat convergentie in waarschijnlijkheid van X_N naar a identiek is met convergentie in verdeling van X_N naar P_a, waarin P_a de kansmaat is die aan het punt a kans 1 toekent. We kunnen $X_N \rightarrow_{P_N} a$ dus ook schrijven als $L(X_N) \rightarrow_z P_a$.

Voortdurend wordt in het vervolg gebruik gemaakt van de volgende bekende eigenschap.

LEMMA 1.8: Laat $\{X_N\}$ en $\{Y_N\}$ rijen stochastische k-vectoren ($k \geq 1$) zijn, waarbij X_N en Y_N gedefinieerd zijn op dezelfde kansruimte (Ω_N, A_N, P_N), $N = 1,2,\ldots$. Onderstel dat $X_N \rightarrow_D X$ en $Y_N \rightarrow_{P_N} a$ ($a \in \mathbb{R}^k$). Dan geldt

$$X_N + Y_N \rightarrow_D X + a \quad \text{en} \quad X_N - Y_N \rightarrow_D X - a.$$

Als k = 1, dan geldt bovendien

$$X_N . Y_N \rightarrow_D aX \quad \text{en} \quad X_N / Y_N \rightarrow_D X/a \quad (\text{mits } a \neq 0).$$

BEWIJS:
Zie b.v. CHUNG [7] p. 85-86. Het resultaat is ook eenvoudig af te leiden uit stelling 2.9.9 in FABIUS & VAN ZWET [8] en het voorgaande lemma 1.4 (met $h(x,y) := x \pm y$, $h(x,y) := xy$ en $h(x,y) := x/y$ respectievelijk). □

Hiermee besluiten we onze uiteenzetting over zwakke convergentie. We introduceren thans het begrip naburigheid van twee rijen kansmaten.

Tenzij anders vermeld gaan we in de rest van dit hoofdstuk steeds van de volgende situatie uit. Voor elke N = 1,2,... is een meetbare ruimte $(\mathcal{X}_N, \mathcal{A}_N)$ gegeven en op elke $(\mathcal{X}_N, \mathcal{A}_N)$ zijn twee kansmaten P_N en Q_N gedefinieerd.

OPMERKING 1.3: In toepassingen stelt $\mathcal{X}_N$ gewoonlijk de steekproefruimte voor corresponderend met een steekproef van N waarnemingen. De kansmaten P_N en Q_N zijn dan kansverdelingen van de stochastische waarnemingsvector X_N (met waardenbereik $\mathcal{X}_N$), waarbij P_N b.v. correspondeert met een punt uit een nulhypothese en Q_N met een punt uit een alternatieve hypothese.

DEFINITIE 1.8: De rijen kansmaten $\{P_N\}$ en $\{Q_n\}$ heten *naburig* (Engels: *contiguous*) als de volgende equivalentie juist is (voor $A_N \in \mathcal{A}_N$, N=1,2,...):

$$(1.6) \qquad \lim_{N\to\infty} P_N(A_N) = 0 \iff \lim_{N\to\infty} Q_N(A_N) = 0.$$

Notatie: $\{P_N\} \triangleleft\triangleright \{Q_N\}$.

De rij $\{Q_N\}$ heet *eenzijdig naburig* t.o.v. $\{P_N\}$ als steeds de volgende implicatie juist is:

$$(1.7) \qquad \lim_{N\to\infty} P_N(A_N) = 0 \Longrightarrow \lim_{N\to\infty} Q_N(A_N) = 0.$$

Notatie: $\{Q_N\} \lhd \{P_N\}$.

OPMERKING 1.4: LECAM ([16],[17]) en ROUSSAS [22] hanteren (1.6) als definitie, terwijl HÁJEK & ŠIDÁK [12] en WITTING & NÖLLE [24] steeds (1.7) als definitie van naburigheid kiezen. Maar in de meeste gevallen waar naburigheid in de zin van (1.7) een rol speelt, is ook (1.6) vervuld. We nemen daarom (1.6) als uitgangspunt van onze beschouwingen.

Formule (1.6) doet denken aan de definitie van absolute continuïteit van kansmaten t.o.v. elkaar. Inderdaad impliceert (1.6) dat de kansmassa van P_N, die singulier is t.o.v. Q_N, voor $N \to \infty$ naar 0 convergeert. Immers zij $\mu_N := P_N + Q_N$ en zij

$$p_N := dP_N/d\mu_N \quad \text{en} \quad q_N := dQ_N/d\mu_N, \quad N = 1,2,\ldots .$$

Zij voorts $A_N := \{x : q_N(x) = 0\}$. Dan is $P_N(A_N)$ de kansmassa van P_N, die singulier is t.o.v. Q_N. Daar $Q_N(A_N) = 0$, volgt uit (1.6) dat $P_N(A_N) \to 0$ voor $N \to \infty$. Evenzo bewijst men dat (1.6) impliceert dat de kansmassa van Q_N, die singulier is t.o.v. P_N, voor $N \to \infty$ naar 0 convergeert. Maar ook al zijn P_N en Q_N voor elke N absoluut continu t.o.v. elkaar, dan behoeven $\{P_N\}$ en $\{Q_N\}$ nog niet naburig te zijn!

OPMERKING 1.5: Is $B \subset \bar{R}$ en h_N een afbeelding van X_N in $\bar{R}$, dan zullen we de verzameling $\{x : h_N(x) \in B\}$ dikwijls kortweg schrijven als $\{h_N \in B\}$.

De volgende stelling geeft een andere formulering van naburigheid van twee rijen kansmaten, die men in de literatuur ook vaak als definitie aantreft in plaats van (1.6).

STELLING 1.9: De volgende twee uitspraken zijn equivalent:

(i) $\{P_N\} \lhd\rhd \{Q_N\}$,

(ii) voor elke rij stochastische grootheden $\{T_N\}$, met T_N gedefinieerd op (X_N,A_N), $N = 1,2,\ldots$, en voor elke $c \in \bar{R}$ geldt dat $T_N \xrightarrow{P_N} c$ d.e.s.d. als $T_N \xrightarrow{Q_N} c$.

BEWIJS:
Zij (i) vervuld en zij $A_N := \{|T_N - c| > \varepsilon\}$, $N = 1,2,\ldots$. Dan volgt uit $P_N(A_N) \to 0$ dat $Q_N(A_N) \to 0$ en omgekeerd (voor elke $\varepsilon>0$), zodat (ii) geldt. Onderstel omgekeerd dat (ii) vervuld is. Als $P_N(A_N) \to 0$ voor zekere rij $\{A_N\}$, kies dan $T_N := \chi(A_N)$, zodat $T_N \to_{P_N} 0$ en dus $T_N \to_{Q_N} 0$, d.w.z. $Q_N(A_N) \to 0$. Als $Q_N(A_N) \to 0$, dan volgt $P_N(A_N) \to 0$ op dezelfde wijze. □

Ook al zijn $\{P_N\}$ en $\{Q_N\}$ naburig, dit betekent nog niet dat P_N en Q_N voor grote N sterk op elkaar gelijken. Wel is het zo dat, als P_N en Q_N voor $N \to \infty$ sterk tot elkaar naderen, dan $\{P_N\}$ en $\{Q_N\}$ naburig zijn. Om dit te laten zien introduceren we een paar afstandsbegrippen tussen kansmaten, en wel in de eerste plaats de variatie-afstand van Kolmogorov.

DEFINITIE 1.9: Zijn P en Q twee kansmaten gedefinieerd op een meetbare ruimte $(\Omega,\mathcal{A})$, dan is de *variatie-afstand* van P en Q gedefinieerd door

$$(1.8) \qquad \|P-Q\| := \sup_{A\in\mathcal{A}} |P(A) - Q(A)|.$$

Men gaat eenvoudig na dat $0 \leq \|P-Q\| \leq 1$ en dat de variatie-afstand een metriek is in de verzameling van alle kansmaten op $(\Omega,\mathcal{A})$. Zijn P en Q absoluut continu t.o.v. een maat μ (b.v. $\mu = P + Q$) dan is de variatie-afstand ook op een andere wijze te schrijven.

LEMMA 1.10: Zijn de kansmaten P en Q op $(\Omega,\mathcal{A})$ absoluut continu t.o.v. een σ-eindige maat μ op $(\Omega,\mathcal{A})$ en is $p := dP/d\mu$ en $q := dQ/d\mu$, dan is

$$(1.9) \qquad \|P-Q\| = \tfrac{1}{2}\int |p-q|\,d\mu.$$

BEWIJS:
Zij $D := \{\omega : p(\omega) - q(\omega) > 0\}$. Dan is

$$\|P-Q\| = \sup_A |P(A) - Q(A)| = \sup_A \left|\int_A (p-q)d\mu\right| =$$

$$= \sup_A \left|\int_{A\cap D} |p-q|d\mu - \int_{A\cap D^c} |p-q|d\mu\right| =$$

$$= \max\left\{\int_D |p-q|d\mu, \int_{D^c} |p-q|d\mu\right\}.$$

Maar omdat p en q dichtheden zijn t.o.v. μ, is

$$\int_D |p-q|d\mu = \int_D (p-q)d\mu = -\int_{D^c} (p-q)d\mu = \int_{D^c} |p-q|d\mu,$$

zodat $\|P-Q\| = \frac{1}{2}\int |p-q|d\mu$. □

Uit dit lemma blijkt direct dat $\|P-Q\| = 1$ d.e.s.d. als $P \perp Q$. Nauw verwant met de variatie-afstand is de Hellinger-afstand.

DEFINITIE 1.10: Zijn P en Q kansmaten op een meetbare ruimte (Ω,A) en zijn P en Q absoluut continu t.o.v. een σ-eindige maat μ (b.v. $\mu = P+Q$) met dichtheden $p := dP/d\mu$ en $q := dQ/d\mu$, dan is de *Hellinger-afstand* van P en Q gedefinieerd door

$$(1.10) \qquad \delta(P,Q) := \left[\int (p^{\frac{1}{2}}-q^{\frac{1}{2}})^2 d\mu\right]^{\frac{1}{2}}$$

De Hellinger-afstand hangt niet af van de keuze van de dominerende maat μ. Immers zij $\mu = P+Q$ en zij ν een andere σ-eindige dominerende maat met $\tilde{p} := dP/d\nu$ en $\tilde{q} := dQ/d\nu$, dan is $\mu << \nu$ en dus

$$\int (p^{\frac{1}{2}}-q^{\frac{1}{2}})^2 d\mu = \int (p^{\frac{1}{2}}-q^{\frac{1}{2}})^2 (d\mu/d\nu)d\nu =$$

$$= \int \{p^{\frac{1}{2}}(d\mu/d\nu)^{\frac{1}{2}} - q^{\frac{1}{2}}(d\mu/d\nu)^{\frac{1}{2}}\}^2 d\nu =$$

$$= \int (\tilde{p}^{\frac{1}{2}}-\tilde{q}^{\frac{1}{2}})^2 d\nu.$$

Merk op dat

$$0 \leq \delta(P,Q) \leq \sqrt{2}$$

met $\delta(P,Q) = 0$ d.e.s.d. als $P = Q$ en $\delta(P,Q) = \sqrt{2}$ d.e.s.d. als $P \perp Q$. De Hellinger-afstand is ook een metriek (afgeleid uit de L_2-norm), die dezelfde metrische topologie in de verzameling van alle kansmaten op (Ω,A) induceert als de variatie-afstand, zoals blijkt uit het volgende lemma.

LEMMA 1.11: Zijn P en Q kansmaten op een meetbare ruimte (Ω,A), dan geldt

$$(1.11) \qquad \tfrac{1}{2}\delta(P,Q)^2 \leq \|P-Q\| \leq \delta(P,Q).$$

BEWIJS:
De eerste ongelijkheid is een direct gevolg van lemma 1.10 en de elementaire ongelijkheid $(p^{\frac{1}{2}}-q^{\frac{1}{2}})^2 \leq |p-q|$. De tweede ongelijkheid bewijzen we door tweemaal de ongelijkheid van Schwarz toe te passen:

$$\begin{aligned}
\|P-Q\| &= \tfrac{1}{2}\int |p-q|\,d\mu = \tfrac{1}{2}\int |(p^{\frac{1}{2}}-q^{\frac{1}{2}})(p^{\frac{1}{2}}+q^{\frac{1}{2}})|\,d\mu \\
&\leq \tfrac{1}{2}\Big[\int (p^{\frac{1}{2}}-q^{\frac{1}{2}})^2 d\mu \int (p^{\frac{1}{2}}+q^{\frac{1}{2}})^2 d\mu\Big]^{\frac{1}{2}} \\
&= \tfrac{1}{2}\delta(P,Q)\Big[2+2\int p^{\frac{1}{2}}q^{\frac{1}{2}}d\mu\Big]^{\frac{1}{2}} \\
&\leq \tfrac{1}{2}\delta(P,Q)[2+2]^{\frac{1}{2}} = \delta(P,Q). \qquad \square
\end{aligned}$$

De laatste ongelijkheid in (1.11) kan verscherpt worden (zie [18]), maar deze verscherping speelt in het vervolg geen rol.

Als $P, Q_1, Q_2, \ldots$ kansmaten zijn op $(\mathbb{R}^k, \mathcal{B}^k)$, dan impliceert $\lim_{n\to\infty} \|P-Q_n\| = 0$ dat $Q_n \to_z P$, zoals direkt blijkt uit definitie 1.9 en lemma 1.3(iv). Men ziet gemakkelijk in dat de omkering van deze bewering niet algemeen juist is.

We keren nu weer terug tot onze rijen kansmaten $\{P_N\}$ en $\{Q_N\}$.

STELLING 1.12: Als $\lim_{N\to\infty}\|P_N-Q_N\| = 0$ of als $\lim_{N\to\infty}\delta(P_N,Q_N) = 0$, dan geldt $\{P_N\} \Leftrightarrow \{Q_N\}$.

BEWIJS:
Onderstel dat $\|P_N-Q_N\| \to 0$, hetgeen volgens lemma 1.11 d.e.s.d. het geval is als $\delta(P_N,Q_N) \to 0$. Zij $\{A_N\}$ een rij verzamelingen zó dat $P_N(A_N) \to 0$. Daar $|P_N(A_N) - Q_N(A_N)| \leq \|P_N-Q_N\|$, volgt hieruit dat ook $Q_N(A_N) \to 0$. Evenzo volgt uit $Q_N(A_N) \to 0$ dat $P_N(A_N) \to 0$, zodat $\{P_N\} \Leftrightarrow \{Q_N\}$. □

In het vervolg zullen we steeds onderstellen dat voor elke N op (X_N,A_N) een σ-eindige maat μ_N gegeven is, die P_N en Q_N domineert (b.v. $\mu_N = P_N+Q_N$). Zij voor N = 1,2,...

$$p_N := dP_N/d\mu_N, \quad q_N := dQ_N/d\mu_N$$

en

$$I_N := \{x \in X_N : p_N(x) > 0\}, \quad J_N := \{x \in X_N : q_N(x) > 0\}.$$

Definieer vervolgens het aannemelijkheidsquotiënt L_N voor elke N door

$$(1.13) \qquad L_N(x) = \begin{cases} q_N(x)/p_N(x) & \text{als } x \in I_N \\ \infty & \text{als } x \in I_N^c. \end{cases}$$

Zij tenslotte Λ_N voor alle N gedefinieerd door log L_N, d.w.z.

$$(1.14)\qquad \Lambda_N(x) = \begin{cases} \log\{q_N(x)/p_N(x)\} & \text{als } x \in I_N \cap J_N \\ \infty & \text{als } x \in I_N^c \\ -\infty & \text{als } x \in I_N \cap J_N^c \end{cases}$$

De defectieve kansverdeling van de stochastische grootheid Λ_N onder de kansmaat P_N respectievelijk Q_N geven we aan met $L(\Lambda_N|P_N)$ respectievelijk $L(\Lambda_N|Q_N)$, $N = 1,2,\ldots$. Uit definitie (1.14) blijkt wel, dat deze defectieve kansverdelingen strikt defectieve kansmaten kunnen zijn.

De volgende fundamentele stelling laat zien dat naburigheid van $\{P_N\}$ en $\{Q_N\}$ gekarakteriseerd kan worden door eenvoudige eigenschappen van de defectieve kansverdelingen van de Λ_N's.

STELLING 1.13: De volgende uitspraken zijn equivalent:

(a) $\{P_N\} \triangleleft\triangleright \{Q_N\}$,

(b) $\{L(\Lambda_N|P_N)\}$ en $\{L(\Lambda_N|Q_N)\}$ zijn beide relatief compact,

(c) $\{L(\Lambda_N|P_N)\}$ is relatief compact en elk zwak limietpunt L voldoet aan $\int \exp(y)\,dL(y) = 1$.

BEWIJS:

Ter vereenvoudiging van de notatie duiden we $L(\Lambda_N|P_N)$ met L_N en $L(\Lambda_N|Q_N)$ met L'_N aan, $N = 1,2,\ldots$.

(b) $\Longrightarrow$ (a). Zij $\{L_N\}$ relatief compact en dus (lemma 1.5) uniform beperkt. Zij $\{A_N\}$ een rij verzamelingen zó dat $Q_N(A_N) \to 0$. We tonen aan dat dan ook $P_N(A_N) \to 0$. Zij $\varepsilon > 0$. Dan bestaan er een $c_\varepsilon > 0$ en een N_ε zó dat $K_N := \{|\Lambda_N| \leq c_\varepsilon\}$ voldoet aan

$$P_N(K_N) = L_N([-c_\varepsilon, c_\varepsilon]) > 1-\varepsilon \qquad \text{voor alle } N \geq N_\varepsilon.$$

Daar $K_N \subset I_N \cap J_N$ en $|\log(q_N/p_N)| \leq c_\varepsilon$ op K_N, d.w.z. $\exp(-c_\varepsilon) \leq p_N/q_N \leq \exp(c_\varepsilon)$ op K_N $(N=1,2,\ldots)$, geldt voor alle $N \geq N_\varepsilon$

$$P_N(A_N) < P_N(A_N \cap K_N) + \varepsilon = \int_{A_N \cap K_N} p_N d\mu_N + \varepsilon =$$

$$= \int_{A_N \cap K_N} p_N/q_N \, dQ_N + \varepsilon \leq$$

$$\leq \exp(c_\varepsilon) \, Q_N(A_N \cap K_N) + \varepsilon \leq$$

$$\leq \exp(c_\varepsilon) \, Q_N(A_N) + \varepsilon,$$

zodat uit $Q_N(A_N) \to 0$ volgt dat $\limsup_{N\to\infty} P_N(A_N) \leq \varepsilon$. Daar $\varepsilon > 0$ willekeurig was, betekent dit dat $P_N(A_N) \to 0$. Uit de relatieve compactheid van $\{L'_N\}$ volgt analoog dat $P_N(A_N) \to 0$ impliceert $Q_N(A_N) \to 0$.

(a) $\Longrightarrow$ (b). Zij $\{P_N\} \Leftrightarrow \{Q_N\}$. Zij $c > 0$ en noem $H_N := \{\Lambda_N > c\} \cap I_N \cap J_N$. Dan is voor alle N

(1.15) $$P_N(\Lambda_N > c) = P_N(H_N) = \int_{H_N} p_N \, d\mu_N = \int_{H_N} p_N/q_N \, dQ_N =$$

$$= \int_{H_N} \exp(-\Lambda_N) dQ_N \leq \exp(-c).$$

Evenzo is, met $G_N := \{\Lambda_N < -c\} \cap I_N \cap J_N$, voor alle N

(1.16) $$Q_N(\Lambda_N < -c) = Q_N(G_N) = \int_{G_N} q_N d\mu_N = \int_{G_N} q_N/p_N \, dP_N =$$

$$= \int_{G_N} \exp(\Lambda_N) dP_N \leq \exp(-c).$$

Kies $\varepsilon > 0$. Dan is er op grond van (1.15) en (1.16) een $c_\varepsilon > 0$ zodat

(1.17) $P_N(\Lambda_N > c_\varepsilon) < \varepsilon$ voor alle N

en

(1.18) $Q_N(\Lambda_N < -c_\varepsilon) < \varepsilon$ voor alle N.

Is $\{c_N\}$ een rij met $0 < c_N \uparrow \infty$, dan geldt op grond van (1.15) dat $P_N(\Lambda_N > c_N) \to 0$ en wegens $\{Q_N\} \lhd \{P_N\}$ impliceert dit $Q_N(\Lambda_N > c_N) \to 0$. Maar dan laat men uit het ongerijmde eenvoudig zien dat er ook een $c'_\varepsilon > 0$ en een N'_ε bestaan zodat

(1.19) $\qquad Q_N(\Lambda_N > c'_\varepsilon) < \varepsilon \qquad$ voor alle $N \geq N'_\varepsilon$.

Evenzo volgt uit (1.16) en $\{P_N\} \lhd \{Q_N\}$ dat er een $c^*_\varepsilon > 0$ en een N^*_ε bestaan zodat

(1.20) $\qquad P_N(\Lambda_N < -c^*_\varepsilon) < \varepsilon \qquad$ voor alle $N \geq N^*_\varepsilon$.

Uit (1.17) en (1.20) leiden we af dat voor alle $N \geq N^*_\varepsilon$

$$L_N([-c^*_\varepsilon, c_\varepsilon]) = P_N(-c^*_\varepsilon \leq \Lambda_N \leq c_\varepsilon) > 1-2\varepsilon$$

en uit (1.18) en (1.19) dat voor alle $N \geq N'_\varepsilon$

$$L'_N([-c_\varepsilon, c'_\varepsilon]) = Q_N(-c_\varepsilon \leq \Lambda_N \leq c'_\varepsilon) > 1-2\varepsilon.$$

Derhalve zijn $\{L_N\}$ en $\{L'_N\}$ beide uniform beperkt en dus (lemma 1.5) relatief compact.

(c) $\Longrightarrow$ (b). Zij $\{L_N\}$ relatief compact. Zij $\{m\} \subset \{N\}$ en $\{r\} \subset \{m\}$ zó dat $L_r \to_z L$ voor zekere kansmaat L. Op grond van (c) voldoet L aan $\int \exp(y)dL(y) = 1$. Zij $h \in C_C(\mathbb{R})$ en zij $K \subset \mathbb{R}$ compact zó dat $h = 0$ op K^c. Dan geldt

$$\int h dL'_r = \int_K h dL'_r = \int_{\Lambda_r^{-1}(K)} h(\Lambda_r) dQ_r =$$

$$= \int_{\Lambda_r^{-1}(K)} h(\Lambda_r)\exp(\Lambda_r) dP_r = \int_K h(y)\exp(y) dL_r(y) \to$$

$$\to \int_K h(y)\exp(y) dL(y) = \int h dL' \qquad \text{(voor } r \to \infty\text{)},$$

waarin L' de kansmaat is gedefinieerd door $dL'/dL(y) = \exp(y)$. Uit lemma 1.2 volgt nu dat $L'_r \to_z L'$, zodat $\{L'_N\}$ relatief compact is.

(b) $\Longrightarrow$ (c). Onderstel dat $\{L_N\}$ en $\{L'_N\}$ relatief compact zijn. Zij $\{L_m\}$ een deelrij van $\{L_N\}$ zó dat $L_m \to_z L$. Zij $\varepsilon > 0$. Daar $\{L'_N\}$ uniform beperkt is (lemma 1.5), bestaat er een compacte $K_\varepsilon \subset \mathbb{R}$ en een N_ε zodat $L'_N(K_\varepsilon) > 1-\varepsilon$ voor alle $N \geq N_\varepsilon$ én $L(\partial K_\varepsilon) = 0$. Bovendien is

$$L'_N(K_\varepsilon) = Q_N(\Lambda_N \in K_\varepsilon) = \int_{\{\Lambda_N \in K_\varepsilon\}} q_N d\mu_N =$$

$$= \int_{\{\Lambda_N \in K_\varepsilon\}} \exp(\Lambda_N) dP_N = \int_{K_\varepsilon} \exp(y) dL_N(y),$$

zodat

$$\int_{K_\varepsilon} \exp(y) dL_N(y) > 1-\varepsilon \qquad \text{voor alle } N \geq N_\varepsilon. \tag{1.21}$$

Op grond van lemma 1.4 (zie (1.4)) impliceert $L_m \to_z L$ dat

$$\lim_{m\to\infty} \int_{K_\varepsilon} \exp(y) dL_m(y) = \int_{K_\varepsilon} \exp(y) dL(y).$$

Uit deze relatie en (1.21) leiden we af dat

$$\int_{K_\varepsilon} \exp(y) dL(y) \geq 1-\varepsilon. \tag{1.22}$$

Zij $\{B_N\}$ een rij begrensde verzamelingen uit $\mathcal{B}$ zó dat $B_N \uparrow \mathbb{R}$ én $L(\partial B_N) = 0$ voor alle N. Dan impliceert de monotone convergentiestelling

$$\lim_{N\to\infty} \int_{B_N} \exp(y) dL(y) = \int \exp(y) dL(y).$$

Maar voor elke N is wegens de tweede uitspraak van lemma 1.4

$$\int_{B_N} \exp(y)dL(y) = \lim_{m\to\infty} \int_{B_N} \exp(y)dL_m(y) \leq$$

$$\leq \limsup_{m\to\infty} \int \exp(y)dL_m(y) =$$

$$= \limsup_{m\to\infty} \int_{\{|\Lambda_m|<\infty\}} \exp(\Lambda_m)dP_m =$$

$$= \limsup_{m\to\infty} Q_m(|\Lambda_m|<\infty) \leq 1,$$

zodat ook $\int \exp(y)dL(y) \leq 1$. Uit deze ongelijkheid en (1.22) (voor elke $\varepsilon>0$) volgt dan $\int \exp(y)dL(y) = 1$. Hiermee is het bewijs voltooid. □

Een zelfde type karakterisering van $\{P_N\} \lhd\rhd \{Q_N\}$ is *niet* te geven in termen van de defectieve kansverdelingen van de L_N's. Dit hangt samen met het feit dat $\{L(L_N|P_N)\}$ áltijd relatief compact is. Immers zij $c > 0$ en $H_N := \{L_N > c\} \cap I_N \cap J_N$, dan is

$$P_N(L_N>c) = P_N(H_N) = \int_{H_N} p_N d\mu_N = \int_{H_N} 1/L_N \, dQ_N \leq c^{-1},$$

zodat $P_N(L_N\in[0,c]) \geq 1-c^{-1}$ voor alle N. Daar $c > 0$ willekeurig is, betekent dit dat $\{L(L_N|P_N)\}$ uniform beperkt en dus (lemma 1.5) relatief compact is.

De defectieve kansverdelingen van L_N blijken echter juist geschikt om eenzijdige naburigheid van $\{Q_N\}$ t.o.v. $\{P_N\}$ te karakteriseren.

STELLING 1.14: De volgende uitspraken zijn equivalent:

(a') $\{Q_N\} \lhd \{P_N\}$,

(b') $\{L(L_N|Q_N)\}$ is relatief compact,

(c') elk zwak limietpunt L van $\{L(L_N|P_N)\}$ voldoet aan $\int y dL(y) = 1$.

BEWIJS:

Het bewijs verloopt analoog aan het bewijs van stelling 1.13, met L_N i.p.v.

Λ_N, en met gebruikmaking van de boven bewezen relatieve compactheid van $\{L(L_N|P_N)\}$. □

OPMERKING 1.6: De equivalenties (a) $\Longleftrightarrow$ (c) in stelling 1.13 en (a') $\Longleftrightarrow$ (c') in stelling 1.14 staan bekend als het "eerste lemma van LeCam". De betekenis van dergelijke karakteriseringen van naburigheid in termen van de defectieve kansverdelingen van Λ_N (en L_N) is duidelijk wanneer men bedenkt dat vele statistische procedures gebaseerd zijn op aannemelijkheidsquotiënten. De asymptotische eigenschappen (voor $N\to\infty$) van deze procedures hangen uiteraard ten nauwste samen met de asymptotische eigenschappen van de defectieve kansverdelingen van de Λ_N's (en L_N's); zie ook opmerking 1.3.

Bij het bewijs van (c) $\Longrightarrow$ (b) in stelling 1.13 hebben we in feite laten zien dat uit $\{P_N\} \lhd\rhd \{Q_N\}$ volgt dat voor deelrijen $\{m\} \subset \{N\}$ de volgende implicatie waar is:

$$(1.23) \qquad L(\Lambda_m|P_m) \to_z L \Longrightarrow L(\Lambda_m|Q_m) \to_z L',$$

waarin L' gedefinieerd is door $dL'/dL(y) = \exp(y)$. Dit resultaat kunnen we verder generaliseren.

STELLING 1.15: Zij $\{P_N\} \lhd\rhd \{Q_N\}$ en zij T_N een stochastische k-vector op $(\mathcal{X}_N,\mathcal{A}_N)$, $N = 1,2,\ldots$. Dan geldt voor deelrijen $\{m\} \subset \{N\}$ de volgende implicatie:

$$L((\Lambda_m,T_m)|P_m) \to_z \tilde{L} \Longrightarrow L((\Lambda_m,T_m)|Q_m) \to_z \tilde{L}',$$

waarin $\tilde{L}'$ de kansmaat op $(\mathbb{R}^{k+1},\mathcal{B}^{k+1})$ is gedefinieerd door

$$\frac{d\tilde{L}'}{d\tilde{L}}(y,t) = \exp(y) \qquad (y\in\mathbb{R},\ t\in\mathbb{R}^k).$$

BEWIJS:
We schrijven $\tilde{L}_m := L((\Lambda_m,T_m)|P_m)$, $\tilde{L}'_m := L((\Lambda_m,T_m)|Q_m)$ en L_m en L'_m als in

het bewijs van stelling 1.13. Onderstel dat $\tilde{L}_m \to_z \tilde{L}$. Zij $L := \tilde{L}^{\pi}$, waarin π de projectie is van $\mathbb{R}^{k+1}$ op de eerste coördinaat. Op grond van de eerste bewering van lemma 1.4 (of op grond van lemma 1.7) geldt dan $L_m \to_z L$. Met behulp van (1.23) leiden we hieruit af dat $L'_m \to_z L'$ met L' gedefinieerd door $dL'/dL(y) = \exp(y)$. Zij $h \in C_C(\mathbb{R}^{k+1})$ en zij $K \subset \mathbb{R}^{k+1}$ de compacte drager van h. Dan is

$$\int h d\tilde{L}'_m = \int_K h d\tilde{L}'_m = \int_{(\Lambda_m,T_m)^{-1}(K)} h(\Lambda_m,T_m) dQ_m =$$

$$= \int_{(\Lambda_m,T_m)^{-1}(K)} h(\Lambda_m,T_m)\exp(\Lambda_m) dP_m =$$

$$= \int_K h(y,t)\exp(y) d\tilde{L}_m(y,t)$$

$$\to \int_K h(y,t)\exp(y) d\tilde{L}(y,t) = \int h(y,t) d\tilde{L}'(y,t),$$

waarin $\tilde{L}'$ een kansmaat is omdat

$$\int_{\mathbb{R}^{k+1}} \exp(y) d\tilde{L}(y,t) = \int_{\mathbb{R}} \exp(y) dL(y) = 1.$$

Maar dan volgt uit lemma 1.2 dat $\tilde{L}'_m \to_z \tilde{L}'$. □

VOORBEELD 1.1: Zij voor N = 1,2,... X_N een Poisson verdeelde stochastische grootheid met parameter λ_N (>0) óf met parameter λ'_N (>0). In het eerste geval geven we de kansverdeling van X_N aan met P_N, in het tweede geval met Q_N (N=1,2,...). In deze situatie kiezen we als dominerende maat μ_N op de ruimte $\mathcal{X}_N$ van de niet-negatieve gehele getallen de telmaat; dan is

$$\Lambda_N = \Lambda_N(X_N) = \log\{\frac{\exp(-\lambda'_N).(\lambda'_N)^{X_N}/X_N!}{\exp(-\lambda_N).(\lambda_N)^{X_N}/X_N!}\} =$$

$$= (\lambda_N-\lambda'_N) + X_N \log(\lambda'_N/\lambda_N), \qquad N = 1,2,\ldots .$$

Onderstel dat de rijen $\{\lambda_N\}$ en $\{\lambda'_N\}$ beiden naar boven begrensd en van nul verwijderd zijn. Dan is $\{P_N\} \Leftrightarrow \{Q_N\}$.

Immers uit de gemaakte onderstellingen volgt dat elke deelrij $\{m\} \subset \{N\}$ een subdeelrij $\{r\}$ bevat zó dat $\lambda_r \to \lambda$ en $\lambda'_r \to \lambda'$ voor zekere positieve getallen λ en λ' (die van de subdeelrij $\{r\}$ kunnen afhangen). Geven we een Poisson-verdeling met parameter λ aan met $\mathcal{P}_\lambda$, dan geldt dus $P_r \to_z \mathcal{P}_\lambda$ en $Q_r \to_z \mathcal{P}_{\lambda'}$. Hieruit volgt dat

$$\Lambda_r \to_D (\lambda-\lambda') + X \log(\lambda'/\lambda) \qquad \text{voor } r \to \infty,$$

waarin de kansverdeling van de stochastische grootheid X in het eerste geval $\mathcal{P}_\lambda$ en in het tweede geval $\mathcal{P}_{\lambda'}$ is. Hiermee is bewezen dat aan eigenschap (b) van stelling 1.13 voldaan is, zodat op grond van deze stelling $\{P_N\} \Leftrightarrow \{Q_N\}$.

We verifiëren nog eigenschap (c) van stelling 1.13. Geven we de limietverdeling van Λ_r in het eerste geval aan met L, dan geldt

$$\int \exp(y)\,dL(y) = E_\lambda \exp\{(\lambda-\lambda') + X \log(\lambda'/\lambda)\} =$$

$$= \exp(\lambda-\lambda') \sum_{x=0}^{\infty} \exp(-\lambda) \frac{\lambda^x}{x!} \exp(x \log(\lambda'/\lambda)) =$$

$$= \exp(-\lambda') \sum_{x=0}^{\infty} \frac{(\lambda')^x}{x!} = 1.$$

VOORBEELD 1.2: Zij voor $N = 1,2,\ldots$ X_N een stochastische grootheid met òf een normale $N(\mu_N;\sigma_N^2)$ verdeling òf een normale $N(\nu_N;\sigma_N^2)$ verdeling. In het eerste geval duiden we de kansverdeling van X_N aan met P_N, in het tweede geval met Q_N $(N=1,2,\ldots)$. Kiezen we voor de dominerende maat steeds de Lebesgue-maat op $(\mathcal{X}_N,\mathcal{A}_N) := (\mathbb{R},\mathcal{B})$, dan vinden we

$$\Lambda_N = \Lambda_N(X_N) = \log\left\{\frac{(2\pi\sigma_N^2)^{-\frac{1}{2}} \exp[-\frac{1}{2}\sigma_N^{-2}(X_N-\nu_N)^2]}{(2\pi\sigma_N^2)^{-\frac{1}{2}} \exp[-\frac{1}{2}\sigma_N^{-2}(X_N-\mu_N)^2]}\right\} =$$

$$= \frac{\nu_N - \mu_N}{\sigma_N^2} X_N + \frac{\mu_N^2 - \nu_N^2}{2\sigma_N^2}, \qquad N = 1,2,\ldots .$$

Onderstel dat de rij $\{(\nu_N-\mu_N)/\sigma_N\}$ begrensd is. Dan is $\{P_N\} \Leftrightarrow \{Q_N\}$.

Immers elke deelrij $\{m\} \subset \{N\}$ bezit dan een subdeelrij $\{r\}$ zodat $(\nu_r-\mu_r)/\sigma_r \to c$ voor $r \to \infty$, waarin $c \in \mathbb{R}$ van de subdeelrij $\{r\}$ kan afhangen. Daar voor alle N

$$L(\Lambda_N|P_N) = N(-\tfrac{1}{2}\frac{(\nu_N-\mu_N)^2}{\sigma_N^2}; \frac{(\nu_N-\mu_N)^2}{\sigma_N^2})$$

en

$$L(\Lambda_N|Q_N) = N(\tfrac{1}{2}\frac{(\nu_N-\mu_N)^2}{\sigma_N^2}; \frac{(\nu_N-\mu_N)^2}{\sigma_N^2}),$$

geldt voor $r \to \infty$ $L(\Lambda_r|P_r) \longrightarrow_z N(-\frac{1}{2}c^2;c^2)$ en $L(\Lambda_r|Q_r) \longrightarrow_z N(\frac{1}{2}c^2;c^2)$. Maar dan zijn $\{L(\Lambda_N|P_N)\}$ en $\{L(\Lambda_N|Q_N)\}$ relatief compact, waaruit met stelling 1.13 de naburigheid van $\{P_N\}$ en $\{Q_N\}$ volgt.

Het is in voorbeeld 1.2 geen toeval, dat de normale limietverdelingen van $\{\Lambda_r\}$ een bijzondere verhouding tussen verwachting en variantie te zien geven. Dit blijkt uit de gevolgen, die we thans bewijzen.

GEVOLG 1.16: Als $L(\Lambda_N|P_N) \longrightarrow_z N(\mu;\sigma^2)$ met $\sigma^2 \geq 0$, dan geldt

$$\{P_N\} \Leftrightarrow \{Q_N\} \qquad \text{d.e.s.d. als } \mu = -\tfrac{1}{2}\sigma^2.$$

BEWIJS:

Onderstel dat $\{P_N\} \Leftrightarrow \{Q_N\}$. Zij $L := N(\mu;\sigma^2)$, dan is volgens stelling 1.13 (zie (c)) $\int \exp(y)dL(y) = 1$. Als $\sigma^2 = 0$, dan geeft dit $\exp(\mu) = 1$, d.w.z. $\mu = 0$. Als $\sigma^2 > 0$, dan is

$$\int \exp(y)dL(y) = (2\pi\sigma^2)^{-\frac{1}{2}} \int \exp\{y-\tfrac{1}{2}\sigma^{-2}(y-\mu)^2\}dy =$$

$$= (2\pi\sigma^2)^{-\frac{1}{2}} \int \exp\{-\tfrac{1}{2}\sigma^{-2}(y-\mu-\sigma^2)^2 + \mu + \tfrac{1}{2}\sigma^2\}dy =$$

$$= \exp(\mu+\tfrac{1}{2}\sigma^2) = 1,$$

zodat $\mu = -\frac{1}{2}\sigma^2$.

Is omgekeerd $\mu = -\frac{1}{2}\sigma^2$, dan volgt uit het bovenstaande dat $\int \exp(y)dL(y) = 1$, zodat $\{P_N\} \triangleleft\triangleright \{Q_N\}$ wegens stelling 1.13. □

GEVOLG 1.17: Als $L(\Lambda_N|P_N) \rightarrow_z N(-\frac{1}{2}\sigma^2;\sigma^2)$, dan geldt ook $L(\Lambda_N|Q_N) \rightarrow_z N(\frac{1}{2}\sigma^2;\sigma^2)$, voor alle $\sigma^2 \geq 0$.

BEWIJS:
Op grond van gevolg 1.16 is $\{P_N\} \triangleleft\triangleright \{Q_N\}$. Wegens (1.23) geldt dan $L(\Lambda_N|Q_N) \rightarrow_z L'$ met $dL'/dL(y) = \exp(y)$. Als $\sigma^2 = 0$, dan volgt de bewering direct uit stelling 1.9. Als $\sigma^2 > 0$, dan rekent men eenvoudig na (zie bewijs van gevolg 1.16 met $\mu = -\frac{1}{2}\sigma^2$) dat $L' = N(\frac{1}{2}\sigma^2;\sigma^2)$. □

GEVOLG 1.18: Zij T_N een stochastische grootheid op (X_N,A_N), $N = 1,2,\ldots$, en zij $\sigma > 0$. Dan gelden de volgende implicaties:

(i) $\left.\begin{array}{l} \{P_N\} \triangleleft\triangleright \{Q_N\} \\ \sigma T_N - \Lambda_N \rightarrow_{P_N} \frac{1}{2}\sigma^2 \end{array}\right\} \Longrightarrow \sigma T_N - \Lambda_N \rightarrow_{Q_N} \frac{1}{2}\sigma^2,$

(ii) $\left.\begin{array}{l} L(\Lambda_N|P_N) \rightarrow_z N(-\frac{1}{2}\sigma^2;\sigma^2) \\ \sigma T_N - \Lambda_N \rightarrow_{P_N} \frac{1}{2}\sigma^2 \end{array}\right\} \Longrightarrow L(T_N|P_N) \rightarrow_z N(0;1),$

(iii) $\left.\begin{array}{l} L(T_N|P_N) \rightarrow_z N(0;1) \\ \sigma T_N - \Lambda_N \rightarrow_{P_N} \frac{1}{2}\sigma^2 \end{array}\right\} \Longrightarrow L(T_N|Q_N) \rightarrow_z N(\sigma;1).$

BEWIJS:
(i). Dit is een direct gevolg van stelling 1.9.
(ii). Daar $T_N = \sigma^{-1}(\sigma T_N - \Lambda_N) + \sigma^{-1}\Lambda_N$, $N = 1,2,\ldots$, volgt de bewering uit het gegeven en lemma 1.8.
(iii). Daar $\Lambda_N = -(\sigma T_N - \Lambda_N) + \sigma T_N$, volgt uit het gegeven en lemma 1.8 direct dat $L(\Lambda_N|P_N) \rightarrow_z N(-\frac{1}{2}\sigma^2;\sigma^2)$. Op grond van gevolg 1.16 geldt dus $\{P_N\} \triangleleft\triangleright \{Q_N\}$. Uit (i) leiden we dan af dat $\sigma T_N - \Lambda_N \rightarrow_{Q_N} \frac{1}{2}\sigma^2$, terwijl gevolg 1.17 impli-

ceert dat $L(\Lambda_N|Q_N) \to_z N(\frac{1}{2}\sigma^2;\sigma^2)$. Combineren we deze beide laatste resultaten, dan vinden we dat $L(T_N|Q_N) \to_z N(\sigma;1)$. □

De drie bovenstaande gevolgen zullen in de volgende hoofdstukken een voorname plaats innemen. Merk op dat de gegevens van zowel gevolg 1.17 als van (ii) en (iii) in gevolg 1.18 steeds impliceren dat $\{P_N\} \triangleleft\triangleright \{Q_N\}$. Het is op deze indirecte wijze dat naburigheid van twee rijen kansmaten in toepassingen vaak een grote rol speelt.

We besluiten dit hoofdstuk met een stelling die bekend staat als het "derde lemma van LeCam" (zie hoofdstuk II voor het "tweede lemma van LeCam").

STELLING 1.19: Zij T_N een stochastische grootheid op (X_N,A_N), $N = 1,2,\ldots$. Als

$$L((\Lambda_N,T_N)|P_N) \to_z N(-\tfrac{1}{2}\sigma^2,a;\sigma^2,\tau^2,c/\sigma\tau)$$

voor zekere $a \in \mathbb{R}$, $\sigma^2 \geq 0$, $\tau^2 \geq 0$ en $|c| \leq \sigma\tau$, dan geldt

$$L(T_N|Q_N) \to_z N(a+c;\tau^2).$$

BEWIJS:
Daar op grond van het gegeven (zie lemma 1.7) $L(\Lambda_N|P_N) \to_z N(-\frac{1}{2}\sigma^2;\sigma^2)$, is $\{P_N\} \triangleleft\triangleright \{Q_N\}$ wegens gevolg 1.16. Uit stelling 1.15 volgt dan

$$L((\Lambda_N,T_N)|Q_N) \to_z \tilde{L}',$$

waarin

$$\frac{d\tilde{L}'}{dN(-\frac{1}{2}\sigma^2,a;\sigma^2,\tau^2,c/\sigma\tau)}(y,t) = \exp(y).$$

Als $\tau^2 = 0$ (en dus $c = 0$), dan impliceert $T_N \to_{P_N} a$ wegens stelling 1.9 dat ook $T_N \to_{Q_N} a$.
Als $\sigma^2 = 0$ (en dus $c = 0$) én $\tau^2 > 0$, dan is $\exp(y) = 1$ b.o. onder $N(-\frac{1}{2}\sigma^2,a;\sigma^2,\tau^2,c/\sigma\tau)$, zodat in dit geval onmiddellijk volgt dat

$L(T_N|Q_N) \to_z N(a,\tau^2)$.
Zij tenslotte $\sigma^2 > 0$ en $\tau^2 > 0$. In verband met het mogelijk singuliere karakter van de covariantiematrix maken we gebruik van karakteristieke functies. Zij $\tilde{\phi}$ de karakteristieke functie van L', dan is

$$\tilde{\phi}(s_1,s_2) = \int \exp(is_1y+is_2t+y)dN(-\tfrac{1}{2}\sigma^2,a;\sigma^2,\tau^2,c/\sigma\tau)(y,t).$$

De karakteristieke functie van de tweede component is dus

$$\tilde{\phi}(0,s_2) = \int \exp(is_2t+y)dN(-\tfrac{1}{2}\sigma^2,a;\sigma^2,\tau^2,c/\sigma\tau)(y,t).$$

We berekenen deze integraal door eerst de integrand te integreren naar t met de voorwaardelijke verdeling gegeven y en vervolgens te integreren naar y met de marginale verdeling van y. Zij $\rho := c/(\sigma\tau)$ en zij L'_y de (eventueel ontaarde) voorwaardelijke verdeling gegeven y, dan is

$$L'_y = N(a+\tfrac{1}{2}c+\sigma^{-2}cy;\tau^2(1-\rho^2)).$$

Dit geeft voor alle $s_2 \in \mathbb{R}$

$$\tilde{\phi}(0,s_2) = \int (2\pi\sigma^2)^{-\frac{1}{2}}\exp\{y - \tfrac{1}{2}\sigma^{-2}(y+\tfrac{1}{2}\sigma^2)^2\}[\int \exp(is_2t)dL'_y(t)]dy =$$

$$= \int (2\pi\sigma^2)^{-\frac{1}{2}}\exp\{y - \tfrac{1}{2}\sigma^{-2}(y+\tfrac{1}{2}\sigma^2)^2\}\exp\{is_2(a+\tfrac{1}{2}c+\sigma^{-2}cy) +$$

$$-\tfrac{1}{2}\tau^2(1-\rho^2)s_2^2\}dy =$$

$$= \exp\{is_2(a+\tfrac{1}{2}c) - \tfrac{1}{2}\tau^2(1-\rho^2)s_2^2\}\int (2\pi\sigma^2)^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}\sigma^{-2}(y-\tfrac{1}{2}\sigma^2-ics_2)^2 +$$

$$-\tfrac{1}{2}\tau^2\rho^2s_2^2 + \tfrac{1}{2}ics_2\}dy =$$

$$= \exp\{is_2(a+c) - \tfrac{1}{2}\tau^2s_2^2\}$$

en dit is juist de karakteristieke functie van de $N(a+c;\tau^2)$ verdeling. □

OPMERKING 1.7: Men kan de uitspraak van stelling 1.19 nog enigszins versterken. Onder dezelfde condities geldt namelijk ook

$$(1.24) \qquad L((\Lambda_N,T_N)|Q_N) \rightarrow_z N(\tfrac{1}{2}\sigma^2,a+c;\sigma^2,\tau^2,c/\sigma\tau).$$

Men gaat dit eenvoudig na door in het bewijs van stelling 1.19 $\tilde{\phi}(s_1,s_2)$ uit te rekenen in plaats van $\tilde{\phi}(0,s_2)$. Zonder veel moeite vindt men dat voor alle $(s_1,s_2) \in \mathbb{R}^2$

$$\tilde{\phi}(s_1,s_2) = \exp\{\tfrac{1}{2}i\sigma^2 s_1 + i(a+c)s_2 - \tfrac{1}{2}\sigma^2 s_1^2 - cs_1s_2 - \tfrac{1}{2}\tau^2 s_2^2\},$$

waaruit (1.24) direct volgt.

BRONVERMELDINGEN

De besproken theorie over vage en zwakke convergentie is grotendeels ontleend aan BILLINGSLEY [6], CHUNG [7] en FABIUS & VAN ZWET [8]. Alleen lemma 1.1 is overgenomen uit HEYER [13]. Zie overigens ook de appendix in ROUSSAS [22].

De definities en eigenschappen van de variatie-afstand van Kolmogorov en van de Hellinger-afstand worden deels in ROUSSAS [22] en deels in LECAM [17] besproken; zie ook LECAM [19] en KRAFT [15].

De behandeling van naburige rijen kansmaten berust voornamelijk op ROUSSAS [22] en WITTING & NÖLLE [24]. De definities van L_N en Λ_N (zie (1.13) en (1.14) zijn overgenomen uit WITTING & NÖLLE [24]. De formulering van stelling 1.13, ontleend aan ROUSSAS [22], is in verband hiermee enigszins gewijzigd en het bewijs van ROUSSAS [22] daardoor sterk vereenvoudigd, ook al berust het op precies dezelfde ideeën. Overigens zijn stelling 1.13 en 1.14, evenals het merendeel van de daarop volgende resultaten, in essentie reeds aanwezig in het werk van LECAM [16]. Stelling 1.15, voorbeeld 1.1 en 1.2, gevolg 1.16 en 1.17 en stelling 1.19 zijn overgenomen uit ROUSSAS [22]. Tenslotte is gevolg 1.18 een iets gewijzigde versie van een resultaat uit WITTING & NÖLLE [24].

Hoofdstuk 2 ASYMPTOTISCHE TOETSINGSTHEORIE

Dit hoofdstuk is gewijd aan de asymptotische toetsingstheorie, waarbij we ons de resultaten van het vorige hoofdstuk ten nutte zullen maken. We veronderstellen in het vervolg dat de kansmaten op de uitkomstenruimte productmaten zijn. Deze veronderstelling maakt de toepassing van de resultaten van hoofdstuk 1 wezenlijk eenvoudiger.

We beginnen dit hoofdstuk met de invoering van asymptotische toetsingsproblemen en de fundamentele definitie van een rij asymptotisch meest onderscheidende toetsen bij toenemende steekproefomvang N, waarbij nulhypothese en alternatieve hypothese (of kortweg alternatief) voor elke N enkelvoudig zijn. Daar bij toenemende N de limiet van het onderscheidingsvermogen van een rij toetsen in een vast alternatief gewoonlijk 1 is, blijkt het zinvol de alternatieven van N te laten afhangen en wel zó dat de bijbehorende rij kansmaten (eenzijdig) naburig is ten opzichte van de rij kansmaten die met de nulhypothese correspondeert. Uiteraard kunnen we ons enkelvoudige nulhypothesen en enkelvoudige alternatieven steeds ingebed denken in samengestelde nulhypothesen en samengestelde alternatieven.

Vervolgens gaan we over tot het belangrijkste probleem van dit hoofdstuk: de constructie van een rij asymptotisch meest onderscheidende toetsen, bij enkelvoudige nulhypothesen en enkelvoudige alternatieven. De hierbij gemaakte onderstellingen worden nog wat nader toegelicht met behulp van de Hellinger-afstand tussen de (geparametrizeerde) kansverdelingen corresponderend met nulhypothese en alternatief.

Tot besluit van dit hoofdstuk besteden we enige aandacht aan het begrip *asymptotische relatieve doeltreffendheid.*

In dit hoofdstuk beperken we ons tot het bijzondere geval dat $(X,A) = (\mathbb{R}^k,\mathcal{B}^k)$, voor zeker vast natuurlijk getal k, waarop een familie kansmaten $\{P_\zeta,\ \zeta \in T\}$ gedefinieerd is zodanig dat $P_\theta \neq P_\tau$ als $\theta \neq \tau$ $(\theta,\tau\epsilon T)$. We noemen $\zeta \in T$ een *parameter.* Voor N = 1,2,... beschouwen we de productruimte (X^N,A^N), een willekeurig element $\zeta(N) = (\zeta(N,1),\ldots,\zeta(N,N)) \in T^N$ en de bijbehorende productmaat $P_{\zeta(N)} = \Pi_{n=1}^N P_{\zeta(N,n)}$ op deze productruimte. Met het oog op meersteekproeven problemen laten we toe dat alle componenten van deze vector van parameters onderling verschillend zijn. Het zal soms de

formulering ten goede komen als we ons deze productmaat geïnduceerd denken door een stochastische vector $X_N = (X_{N1},\ldots,X_{NN})$, waarvan de componenten X_{Nn} o.o. zijn en waardenbereik in X hebben (n=1,2,...,N). Daarom zullen we (X^N,A^N) ook wel de steekproefruimte noemen. Voor de verwachting van een stochastische grootheid $T = T(X_N)$ zullen wij de notatie $E_{\zeta(N)}T := \int T dP_{\zeta(N)}$ en voor de variantie $Var_{\zeta(N)}T$ gebruiken. Tenslotte zullen wij de verzameling van alle rijen $\{S(N)\}$ van geordende N-tallen parameters met S aanduiden.

Laten nu $\{\theta(N)\}$ en $\{\tau(N)\}$ twee elementen van S zijn. Voor iedere $N = 1,2,\ldots$ en $n = 1,2,\ldots,N$ zij $\mu_{N,n}$ een σ-eindige maat op (X,A), die $P_{\theta(N,n)}$ en $P_{\tau(N,n)}$ domineert (bijvoorbeeld $\mu_{N,n} = P_{\theta(N,n)} + P_{\tau(N,n)}$). Ten opzichte van $\mu_{N,n}$ heeft $P_{\theta(N,n)}$ de dichtheid $g_{\theta(N,n)}$ en $P_{\tau(N,n)}$ de dichtheid $g_{\tau(N,n)}$. Dan zijn de productmaten $P_{\theta(N)}$ en $P_{\tau(N)}$ absoluut continu ten opzichte van $\mu_N = \Pi_{n=1}^N \mu_{N,n}$ met dichtheden

$$p_{\theta(N)}(x_1,\ldots,x_N) = \Pi_{n=1}^N g_{\theta(N,n)}(x_n),$$

(2.1)

$$p_{\tau(N)}(x_1,\ldots,x_N) = \Pi_{n=1}^N g_{\tau(N,n)}(x_n),$$

respectievelijk, voor alle $(x_1,\ldots,x_N) \in X^N$.

Zij $0 \leq \alpha_N \leq 1$ en $\theta(N) \neq \tau(N)$ voor iedere N. Onder het *(eindige) toetsingsprobleem* $(\alpha_N;\theta(N),\tau(N))$ verstaan we het probleem om $P_{\theta(N)}$ tegen $P_{\tau(N)}$ te toetsen met een onbetrouwbaarheidsdrempel α_N. We noemen een toets ψ_N een *toets voor dit probleem*, indien

$$\int \psi_N dP_{\theta(N)} \leq \alpha_N. \tag{2.2}$$

Zij $0 \leq \alpha \leq 1$. Onder het *asymptotische toetsingsprobleem* $(\alpha;\{\theta(N)\},\{\tau(N)\})$ verstaan we het probleem om, voor iedere N, $P_{\theta(N)}$ tegen $P_{\tau(N)}$ te toetsen met een zodanige onbetrouwbaarheidsdrempel α_N dat $\limsup_{N\to\infty} \alpha_N \leq \alpha$. We noemen α de *asymptotische onbetrouwbaarheidsdrempel* en $\{\psi_N\}$ een *rij toetsen voor dit probleem*, indien

$$\limsup_{N\to\infty} \int \psi_N dP_{\theta(N)} \leq \alpha. \tag{2.3}$$

Als de limieten

$$\lim_{N\to\infty} \int \psi_N dP_{\theta(N)} \quad \text{en} \quad \lim_{N\to\infty} \int \psi_N dP_{\tau(N)}$$

bestaan, zullen wij deze respectievelijk de *asymptotische onbetrouwbaarheid* (met betrekking tot $\{P_{\theta(N)}\}$) en het *asymptotisch onderscheidingsvermogen* (met betrekking tot $\{P_{\tau(N)}\}$) van de rij toetsen $\{\psi_N\}$ noemen.

DEFINITIE 2.1: Zij $\{\psi_N\}$ een rij toetsen voor het asymptotische toetsingsprobleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ met $0 \leq \alpha < 1$. Dan heet de rij toetsen $\{\psi_N\}$ *consistent* met betrekking tot de rij alternatieven $\{\tau(N)\}$ als

$$\lim_{N\to\infty} \int \psi_N dP_{\tau(N)} = 1 \tag{2.4}$$

DEFINITIE 2.2. Zij $\{\phi_N\}$ een rij toetsen voor het asymptotische toetsingsprobleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ met $0 \leq \alpha < 1$. Indien

$$\liminf_{N\to\infty} \int (\phi_N - \psi_N) dP_{\tau(N)} \geq 0$$

voor iedere rij toetsen $\{\psi_N\}$ voor het bovengenoemde asymptotische toetsingsprobleem, dan noemt men $\{\phi_N\}$ een *asymptotisch meest onderscheidende* (Engels: *asymptotically most powerful*, afgekort tot *AMP*) *rij toetsen* voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$.

Op grond van het lemma van Neyman-Pearson (zie inleiding) bestaat er een meest onderscheidende toets voor ieder probleem $(\alpha;\theta(N),\tau(N))$ met $0 < \alpha < 1$. Hierop is de volgende stelling gebaseerd.

STELLING 2.1: Zij $0 < \alpha < 1$ en ϕ_N^* de meest onderscheidende toets voor het probleem $(\alpha;\theta(N),\tau(N))$, $N = 1,2,\ldots$. Dan is de rij $\{\phi_N^*\}$ asymptotisch meest onderscheidend voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$.

BEWIJS:
Beschouw een willekeurige rij toetsen $\{\psi_N\}$ voor het probleem

$(\alpha;\{\theta(N)\},\{\tau(N)\})$ en zij $\alpha_N := \int \psi_N dP_{\theta(N)}$. Dan geldt dus dat $\limsup_{N\to\infty} \alpha_N \leq \alpha$. Zij $\tilde{\alpha}_N := \max(\alpha,\alpha_N)$, dan geldt dus $\lim_{N\to\infty} \tilde{\alpha}_N = \alpha$. Zij voorts $\tilde{\phi}_N^*$ de meest onderscheidende toets voor het probleem $(\tilde{\alpha}_N;\theta(N),\tau(N))$, $N = 1,2,\ldots$. Dan is $\int \psi_N dP_{\tau(N)} \leq \int \tilde{\phi}_N^* dP_{\tau(N)}$ voor alle N, zodat de stelling bewezen is zodra we hebben aangetoond dat

$$\lim_{N\to\infty} \int (\phi_N^* - \tilde{\phi}_N^*) dP_{\tau(N)} = 0. \tag{2.5}$$

Zij voor $N = 1,2,\ldots$

$$L_N(x_1,\ldots,x_N) = \begin{cases} p_{\tau(N)}(x_1,\ldots,x_N)/p_{\theta(N)}(x_1,\ldots,x_N) & \text{als } (x_1,\ldots,x_N)\in I_N \\ \infty & \text{als } (x_1,\ldots,x_N)\in I_N^c, \end{cases}$$

waarin $I_N = \{p_{\theta(N)} > 0\}$ (vergelijk (1.13)). Dan hebben op grond van het lemma van Neyman-Pearson ϕ_N^* en $\tilde{\phi}_N^*$ respectievelijk de gedaanten

$$\phi_N^* = \chi_{(c_N,\infty)}(L_N) + \gamma_N \chi_{\{c_N\}}(L_N)$$

en

$$\tilde{\phi}_N^* = \chi_{(\tilde{c}_N,\infty)}(L_N) + \tilde{\gamma}_N \chi_{\{\tilde{c}_N\}}(L_N)$$

voor $N = 1,2,\ldots$. Hierin is $\tilde{c}_N \leq c_N$ voor iedere N daar $\tilde{\alpha}_N \geq \alpha$. Bovendien is $E_{\theta(N)} L_N \leq 1$, zodat wegens de Markov-ongelijkheid

$$\alpha \leq P_{\theta(N)}(L_N \geq c_N) \leq c_N^{-1} E_{\theta(N)} L_N \leq c_N^{-1} \qquad (N=1,2,\ldots).$$

Hieruit volgt $\{\tilde{\phi}_N^* - \phi_N^* > 0\} \subset \{\tilde{c}_N \leq L_N \leq c_N\} \subset \{L_N \leq \alpha^{-1}\}$ voor iedere N. Maar dan is

$$0 \leq \int(\tilde{\phi}_N^*-\phi_N^*)dP_{\tau(N)} = \int_{\{L_N\leq\alpha^{-1}\}}(\tilde{\phi}_N^*-\phi_N^*)dP_{\tau(N)} =$$

$$= \int_{I_N\cap\{p_{\tau(N)}/p_{\theta(N)}\leq\alpha^{-1}\}}(\tilde{\phi}_N^*-\phi_N^*)p_{\tau(N)}d\mu_N \leq$$

$$\leq \alpha^{-1}\int_{\{L_N\leq\alpha^{-1}\}}(\tilde{\phi}_N^*-\phi_N^*)p_{\theta(N)}d\mu_N \leq$$

$$\leq \alpha^{-1}\int(\tilde{\phi}_N^*-\phi_N^*)dP_{\theta(N)} = \alpha^{-1}(\tilde{\alpha}_N-\alpha) \to 0,$$

zodat (2.5) inderdaad vervuld is. □

Aan de hand van het volgende voorbeeld zullen we zien dat het asymptotisch onderscheidingsvermogen de waarden 1 en α, maar ook een willekeurige waarde tussen α en 1 kan aannemen. Dit laatste geval is voor ons doel het meest interessant. Deze drie situaties doen zich voor al naar gelang de "afstand" tussen nulhypothese en alternatief te groot blijft, te snel afneemt, of op een geschikte wijze convergeert.

VOORBEELD 2.1: Laten de componenten X_{Nn} van de stochastische vector X_N hun waardenbereik in $\mathbb{R}$ hebben en laten zij een $N(\zeta(N,n);1)$ verdeling bezitten. Een voorbeeld van een één-steekproef probleem is het probleem $(\alpha;\theta(N),\tau(N))$, waarin $0 < \alpha < 1$, $\theta(N,n) = \theta_0$ en $\tau(N,n) = \theta_N > \theta_0$ voor vaste θ_0 en $n = 1,2,\ldots,N$. Zij $\bar{X}_{N.} = N^{-1}\sum_{n=1}^{N} X_{Nn}$. Voor iedere N is dan de toets $\phi_N^* := \chi(\{N^{\frac{1}{2}}(\bar{X}_{N.}-\theta_0)>\xi_{1-\alpha}\})$ meest onderscheidend voor dit probleem. Op grond van stelling 2.1 is derhalve de rij toetsen $\{\phi_N^*\}$ asymptotisch meest onderscheidend voor ieder probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$, waarin $\theta(N)$ en $\tau(N)$ voor iedere N van het zojuist beschreven type zijn.

Het is eenvoudig in te zien dat

$$\int\phi_N^*dP_{\tau(N)} = P_{\tau(N)}(N^{\frac{1}{2}}(\bar{X}_{N.}-\theta_N)>\xi_{1-\alpha}-N^{\frac{1}{2}}(\theta_N-\theta_0)) =$$

$$= 1 - \Phi(\xi_{1-\alpha}-N^{\frac{1}{2}}(\theta_N-\theta_0)).$$

Hieruit volgt dat

$$\int \phi_N^* dP_{\tau(N)} \to \begin{cases} 1 & \text{als } N^{\frac{1}{2}}(\theta_N - \theta_0) \to +\infty \\ 1 - \Phi(\xi_{1-\alpha} - \eta) & \text{als } N^{\frac{1}{2}}(\theta_N - \theta_0) \to \eta > 0 \\ \alpha & \text{als } N^{\frac{1}{2}}(\theta_N - \theta_0) \to 0. \end{cases}$$

Als $N^{\frac{1}{2}}(\theta_N - \theta_0) \to \infty$, dan is de rij toetsen $\{\phi_N^*\}$ dus consistent. Voor de asymptotische theorie zijn echter rijen $\{\tau(N)\}$ met $\tau(N,n) = \theta_0 + \eta N^{-\frac{1}{2}} + o(N^{-\frac{1}{2}})$ voor $N \to \infty$ het meest van belang.

VOORBEELD 2.2: Laten de componenten X_{Nn} van de stochastische vector X_N als waardenbereik $\mathbb{R}^2$ hebben en een twee-dimensionale normale $N(0,0;1,1,\zeta(N,n))$ verdeling bezitten met $-1 < \zeta(N,n) < 1$ (alle N en n). De limiet van het onderscheidingsvermogen in $\tau(N)$ van de meest onderscheidende toets van de nulhypothese $\theta(N,n) = 0$ (alle N en n) tegen $\tau(N)$ ligt weer strikt tussen α en 1 bij de keuze

$$\tau(N,n) = \eta N^{-\frac{1}{2}} + o(N^{-\frac{1}{2}}), \qquad n = 1,2,\ldots,N,$$

voor vaste doch willekeurige $\eta \neq 0$, als $N \to \infty$.

VOORBEELD 2.3: In de notatie van voorbeeld 2.1 verkrijgen wij een ander voorbeeld door de nulhypothese $\theta(N,n) = \theta_0$ voor $n = 1,\ldots,N$ te toetsen tegen het alternatief $\tau(N,n) = \tau_{1N}$ voor $n = 1,\ldots,N_1$ en $\tau(N,n) = \tau_{2N}$ voor $n = N_1+1,\ldots,N$. Hierin is $N = N_1 + N_2$, θ_0 een vaste constante en $\tau_{1N} \neq \tau_{2N}$. De limiet van het onderscheidingsvermogen in $\tau(N)$ van de meest onderscheidende toets ligt strikt tussen α en 1 bij de keuze

$$\tau(N,n) = \theta_0 - \eta N_1^{-\frac{1}{2}} N_2^{\frac{1}{2}} N^{-\frac{1}{2}} + o(N_1^{-\frac{1}{2}}), \qquad n = 1,\ldots,N_1,$$

$$\tau(N,n) = \theta_0 + \eta N_1^{\frac{1}{2}} N_2^{-\frac{1}{2}} N^{-\frac{1}{2}} + o(N_2^{-\frac{1}{2}}), \qquad n = N_1+1,\ldots,N,$$

voor willekeurige doch vaste $\eta \neq 0$, als $N, N_1, N_2 \to \infty$.

In de volgende stelling wordt een klasse van asymptotisch meest onderscheidende toetsen aangegeven, onder voorwaarden die naburigheid impliceren. Het bewijs van de stelling berust op enkele algemene resultaten van het vorige hoofdstuk. In de definitie van Λ_N in (1.14) moet men nu p_N door $p_{\theta(N)}$ en q_N door $p_{\tau(N)}$ vervangen denken.

STELLING 2.2: Stel dat voor zekere constante $0 < \sigma < \infty$ geldt

$$L(\Lambda_N|P_{\theta(N)}) \longrightarrow_z N(-\tfrac{1}{2}\sigma^2;\sigma^2),$$

als $N \to \infty$, en laat T_N een stochastische grootheid zijn op (X^N,A^N), zodat

$$\sigma T_N - \Lambda_N \longrightarrow_{P_{\theta(N)}} \tfrac{1}{2}\sigma^2,$$

als $N \to \infty$. Dan is $\{P_{\theta(N)}\} \lhd\rhd \{P_{\tau(N)}\}$ en, als $N \to \infty$, geldt

(2.6) $$\sigma T_N - \Lambda_N \longrightarrow_{P_{\tau(N)}} \tfrac{1}{2}\sigma^2,$$

(2.7) $$L(T_N|P_{\theta(N)}) \longrightarrow_z N(0;1), \quad L(T_N|P_{\tau(N)}) \longrightarrow_z N(\sigma;1).$$

Kies $\{c_N\}$ zodanig dat $c_N \to \xi_{1-\alpha}$, als $N \to \infty$, voor zekere $0 < \alpha < 1$. Dan is bovendien de rij toetsen $\{\phi_N\} := \{\chi(\{T_N>c_N\})\}$ asymptotisch meest onderscheidend voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$, met asymptotisch onderscheidingsvermogen

(2.8) $$\lim_{N\to\infty} \int \phi_N dP_{\tau(N)} = 1 - \Phi(\xi_{1-\alpha}-\sigma).$$

BEWIJS:

De naburigheid volgt onmiddellijk uit het gegeven en gevolg 1.16. De relaties (2.6) en (2.7) leidt men direct af uit gevolg 1.18.

Uit het gegeven blijkt dat $\{\phi_N\}$ een rij toetsen voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ is, aangezien $\lim_{N\to\infty} P_{\theta(N)}(T_N>c_N) = \alpha$. Op grond van het lemma van Neyman-Pearson kan men ook $(\Lambda_N + \frac{1}{2}\sigma^2)/\sigma$ als toetsingsgrootheid

voor de meest onderscheidende toets voor het probleem $(\alpha;\theta(N),\tau(N))$ gebruiken. Dan is dus op grond van stelling 2.1 de rij toetsen

$$\{\phi_N^*\} := \{\chi(\{(\Lambda_N+\tfrac{1}{2}\sigma^2)/\sigma>k_N\})\}$$

voor geschikt gekozen k_N asymptotisch meest onderscheidend voor $(\alpha;\{\theta(N)\},\{\tau(N)\})$. Uit het gegeven van de stelling volgt dat $k_N \to \xi_{1-\alpha}$ als $N \to \infty$. Wegens (2.6) en (2.7) geldt onder het alternatief

$$\lim_{N\to\infty} P_{\tau(N)}((\Lambda_N+\tfrac{1}{2}\sigma^2)/\sigma>k_N) =$$

$$= \lim_{N\to\infty} P_{\tau(N)}((\Lambda_N-\tfrac{1}{2}\sigma^2)/\sigma>k_N-\sigma) =$$

$$= 1 - \Phi(\xi_{1-\alpha}-\sigma) =$$

$$= \lim_{N\to\infty} P_{\tau(N)}(T_N>c_N),$$

waarin de tweede overgang geïmpliceerd wordt door gevolg 1.17. Maar dit betekent dat ook de rij toetsen $\{\phi_N\}$ asymptotisch meest onderscheidend is voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$. □

Ook de volgende stelling sluit bij de resultaten van het vorige hoofdstuk aan. Omdat wij nu met productmaten te maken hebben kunnen wij voor iedere $(x_1,\ldots,x_N) \in I_N \cap J_N$ (zie (1.14)) schrijven

$$\Lambda_N = \log[p_{\tau(N)}/p_{\theta(N)}] = \sum_{n=1}^{N} \log[g_{\tau(N,n)}/g_{\theta(N,n)}] = \tag{2.9}$$

$$= 2\sum_{n=1}^{N} \log[g_{\tau(N,n)}/g_{\theta(N,n)}]^{\frac{1}{2}} =$$

$$= 2\sum_{n=1}^{N} ([g_{\tau(N,n)}/g_{\theta(N,n)}]^{\frac{1}{2}}-1) + \text{restterm.}$$

Met het oog op (2.9) is het nuttig de stochastische grootheden

$$(2.10)\qquad W_{Nn} = \begin{cases} 2([g_{\tau(N,n)}(X_{Nn})/g_{\theta(N,n)}(X_{Nn})]^{\frac{1}{2}}-1) & \text{als } X_{Nn} \in I_{Nn}, \\ 0 & \text{anders,} \end{cases}$$

$$W_N = \sum_{n=1}^{N} W_{Nn},$$

in te voeren. De bijzondere vorm waarin Λ_N in het geval van productmaten gegeven kan worden, maakt de verificatie van de volgende hulpstelling met behulp van de centrale limietstelling mogelijk. Deze hulpstelling staat bekend als het "tweede lemma van LeCam".

STELLING 2.3: Laat voor iedere $\varepsilon > 0$ de voorwaarde

$$(2.11)\qquad \lim_{N\to\infty} \max_{n=1,\ldots,N} P_{\theta(N)}(|g_{\tau(N,n)}(X_{Nn})/g_{\theta(N,n)}(X_{Nn})-1|>\varepsilon) = 0,$$

vervuld zijn.

Dan geldt, als $N \to \infty$, voor iedere $0 \leq \sigma < \infty$ de volgende implicatie:

$$(2.12)\qquad L(W_N|P_{\theta(N)}) \longrightarrow_z N(-\sigma^2/4,\sigma^2) \Longrightarrow \begin{cases} W_N - \Lambda_N \longrightarrow_{P_{\theta(N)}} \sigma^2/4, \\ L(\Lambda_N|P_{\theta(N)}) \longrightarrow_z N(-\frac{1}{2}\sigma^2,\sigma^2). \end{cases}$$

BEWIJS:

Onderstel dat de relatie links in (2.12) juist is. Het is voldoende te bewijzen dat

$$W_N - \Lambda_N \longrightarrow_{P_{\theta(N)}} \sigma^2/4,$$

omdat de andere bewering hier een onmiddellijk gevolg van is.

Zij $h : \mathbb{R} \to \mathbb{R}$ tweemaal continu differentieerbaar. Dan is voor elke $u_0 \in \mathbb{R}$

$$h(u) = h(u_0) + (u-u_0)h'(u_0) + \tfrac{1}{2}(u-u_0)^2 \int_0^1 2(1-z)h''(u_0+z(u-u_0))dz,$$

zoals men eenvoudig inziet door partiële integratie van de integraal. Hieruit volgt dat voor $X_N \in I_N \cap J_N$

$$\log(g_{\tau(N,n)}(X_{Nn})/g_{\theta(N,n)}(X_{Nn})) = 2\log(1+\tfrac{1}{2}W_{Nn}) =$$

$$= W_{Nn} - (\tfrac{1}{4})W_{Nn}^2 \int_0^1 2(1-z)(1+\tfrac{1}{2}zW_{Nn})^{-2}dz,$$

en dus

$$\Lambda_N = W_N - (\tfrac{1}{4}) \sum_{n=1}^N W_{Nn}^2 \int_0^1 2(1-z)(1+\tfrac{1}{2}zW_{Nn})^{-2}dz.$$

Dit geldt zelfs voor alle $X_N \in I_N$ (als $X_N \in I_N \cap J_N^c$, dan zijn beide leden $-\infty$), zodat de gelijkheid $P_{\theta(N)}$ - b.o. geldt.

Definieer voor $\delta > 0$

$$W_{Nn,\delta} := \begin{cases} W_{Nn} & \text{als } |W_{Nn}| \leq \delta, \\ 0 & \text{anders,} \end{cases}$$

$n = 1,\ldots,N$ en $N = 1,2,\ldots$. Op grond van de normale convergentiestelling (zie LOÈVE [21], p.316) impliceert voorwaarde (2.11) en $L(W_N|P_{\theta(N)}) \longrightarrow_z N(-\sigma^2/4;\sigma^2)$ dat voor elke $\varepsilon > 0$ en elke $\delta > 0$

$$\lim_{N\to\infty} \sum_{n=1}^N P_{\theta(N)}(|W_{Nn}|>\varepsilon) = 0, \tag{2.13}$$

$$\lim_{N\to\infty} \sum_{n=1}^N E_{\theta(N)}W_{Nn,\delta} = -\sigma^2/4, \tag{2.14}$$

$$\lim_{N\to\infty} \sum_{n=1}^N Var_{\theta(N)}W_{Nn,\delta} = \sigma^2. \tag{2.15}$$

Uit (2.13) volgen de beide relaties

$$\max_{1\leq n\leq N} \left|\int_0^1 2(1-z)(1+\tfrac{1}{2}zW_{Nn})^{-2}dz - 1\right| \longrightarrow_{P_{\theta(N)}} 0,$$

$$\lim_{N\to\infty} P_{\theta(N)}(\sum_{n=1}^{N} W_{Nn}^2 = \sum_{n=1}^{N} W_{Nn,\delta}^2) = 1.$$

Hieruit volgt dat voor $N \to \infty$

$$\sum_{n=1}^{N} W_{Nn}^2 \int_0^1 2(1-z)(1+\tfrac{1}{2}zW_{Nn})^{-2}dz \left[\sum_{n=1}^{N} W_{Nn}^2\right]^{-1} \to_{P_{\theta(N)}} 1.$$

Het is dus voldoende te bewijzen dat

$$\sum_{n=1}^{N} W_{Nn}^2 \to_{P_{\theta(N)}} \sigma^2.$$

We zullen nu laten zien dat voor elke vaste $\varepsilon > 0$ en $\eta > 0$ er een $\delta > 0$ bestaat zodat

$$P(|\sum_{n=1}^{N} W_{Nn,\delta}^2 - \sigma^2| > \varepsilon) < \eta$$

als N voldoende groot. Tezamen met (2.13) impliceert dit juist dat $\Sigma_{n=1}^{N} W_{Nn}^2$ in waarschijnlijkheid convergeert naar σ^2.

Hiertoe bewijzen wij eerst dat

$$\lim_{N\to\infty} \sum_{n=1}^{N} E_{\theta(N)} W_{Nn,\delta}^2 = \sigma^2. \tag{2.16}$$

Daar (2.15) gegeven is, is (2.16) equivalent met

$$\lim_{N\to\infty} \sum_{n=1}^{N} (E_{\theta(N)} W_{Nn,\delta})^2 = 0. \tag{2.17}$$

Zij $\delta > 2$. Dan is $W_{Nn,\delta} \leq W_{Nn}$, omdat $W_{Nn} \geq -2$ op grond van de definitie van W_{Nn}. Toepassing van de ongelijkheid van Schwarz geeft

$$E_{\theta(N)} W_{Nn,\delta} \leq E_{\theta(N)} W_{Nn} = 2\int_{I_{Nn}} [(g_{\tau(N,n)}/g_{\theta(N,n)})^{\frac{1}{2}} - 1]dP_{\theta(N,n)} =$$

$$= 2\int_{I_{Nn}} g_{\tau(N,n)}^{\frac{1}{2}} g_{\theta(N,n)}^{\frac{1}{2}} d\mu_{N,n} - 2P_{\theta(N,n)}(I_{Nn}) \leq$$

$$\leq 2[\int_{I_{Nn}} g_{\tau(N,n)} d\mu_{N,n} \times \int_{I_{Nn}} g_{\theta(N,n)} d\mu_{N,n}]^{\frac{1}{2}} - 2 =$$

$$= 2[P_{\tau(N,n)}(I_{Nn})]^{\frac{1}{2}} - 2 \leq 0.$$

Maar dan is voor $\delta > 2$

$$\sum_{n=1}^{N}(E_{\theta(N)}W_{Nn,\delta})^2 \leq [\min_{1\leq n\leq N}E_{\theta(N)}W_{Nn,\delta}] \sum_{n=1}^{N} E_{\theta(N)}W_{Nn,\delta}.$$

Voorts volgt uit voorwaarde (2.13) en de begrensheid van $W_{Nn,\delta}$ dat

$$\lim_{N\to\infty} \min_{1\leq n\leq N} E_{\theta(N)} W_{Nn,\delta} = 0.$$

Mede op grond van (2.14) vinden wij dus dat

$$\lim_{N\to\infty} \sum_{n=1}^{N} (E_{\theta(N)} W_{Nn,\delta})^2 = 0,$$

voor $\delta > 2$. Maar dan geldt deze relatie voor elke $\delta > 0$. Immers, als $\gamma \leq 2 < \delta$, dan is

$$\sum_{n=1}^{N}(E_{\theta(N)}W_{Nn,\gamma})^2 = \sum_{n=1}^{N} E_{\theta(N)}W_{Nn,\gamma}^2 - \sum_{n=1}^{N} Var_{\theta(N)}W_{Nn,\gamma} \leq$$

$$\leq \sum_{n=1}^{N} E_{\theta(N)}W_{Nn,\delta}^2 - \sum_{n=1}^{N} Var_{\theta(N)}W_{Nn,\gamma} =$$

$$= \sum_{n=1}^{N}(E_{\theta(N)}W_{Nn,\delta})^2 + \sum_{n=1}^{N} Var_{\theta(N)}W_{Nn,\delta} - \sum_{n=1}^{N} Var_{\theta(N)}W_{Nn,\gamma} \to$$

$$\to 0 + \sigma^2 - \sigma^2 = 0,$$

wegens (2.15). Hiermee is (2.17) en dus (2.16) bewezen.

Vervolgens merken wij op dat voor elke $\delta > 0$

$$\sum_{n=1}^{N} Var_{\theta(N)} W_{Nn,\delta}^2 \leq \sum_{n=1}^{N} E_{\theta(N)} W_{Nn,\delta}^4 \leq \delta^2 \sum_{n=1}^{N} E_{\theta(N)} W_{Nn,\delta}^2,$$

zodat op grond van (2.16)

$$\limsup_{N\to\infty} \sum_{n=1}^{N} Var_{\theta(N)} W_{Nn,\delta}^2 \leq \delta^2\sigma^2.$$

Combineren wij dit met (2.16), dan volgt uit de ongelijkheid van Chebyshev dat bij vaste $\varepsilon > 0$ en $\eta > 0$

$$P_{\theta(N)}(|\sum_{n=1}^{N} W_{Nn,\delta}^2 - \sigma^2| > \varepsilon) < \eta$$

voor voldoend kleine δ en alle $N \geq N_\delta$, waarmee het bewijs voltooid is. □

De voorwaarde (2.11), die inhoudt dat de W_{Nn} uniform asymptotisch verwaarloosbaar zijn, is de zwakste algemeen bekende voorwaarde waaronder nog centrale limietstellingen (zonder eisen omtrent het bestaan van momenten) bewezen kunnen worden. In WITTING & NÖLLE [24] wordt in plaats van (2.11) de sterkere voorwaarde opgelegd dat

$$\lim_{N\to\infty} E_{\theta(N)} W_N = -\sigma^2/4 \quad \text{en} \quad \lim_{N\to\infty} Var_{\theta(N)} W_N = \sigma^2,$$

waardoor het bewijs iets eenvoudiger wordt.

Als men in stelling 2.3 eist dat $L(W_N \mid P_{\theta(N)}) \to_z N(\mu;\sigma^2)$ met $\mu \neq -\sigma^2/4$, dan vindt men op precies dezelfde wijze dat $L(\Lambda_N \mid P_{\theta(N)}) \to_z N(\mu-\sigma^2/4;\sigma^2)$; dan zijn $\{P_{\theta(N)}\}$ en $\{P_{\tau(N)}\}$ dus niet naburig (zie gevolg 1.16).

Alvorens wij de centrale stelling van dit hoofdstuk, stelling 2.5, formuleren, vermelden wij nog een technisch lemma.

LEMMA 2.4: Zij $(\Omega,\mathcal{A})$ een meetbare ruimte, voorzien van een σ-eindige maat μ. Zij $\{h_N\}$ een rij kwadratisch μ-sommeerbare functies, welke μ-b.o. convergeert naar een kwadratisch μ-sommeerbare functie h. Als

$$\limsup_{N\to\infty} \int h_N^2 d\mu \leq \int h^2 d\mu,$$

dan geldt

$$\lim_{N\to\infty} \int (h_N - h)^2 d\mu = 0.$$

BEWIJS:
Zie HÁJEK & ŠIDÁK [12], p.154 of, voor een soortgelijk resultaat, WITTING & NÖLLE [24], p.178. □

Aangezien de voorwaarde, waaronder wij stelling 2.5 zullen bewijzen, ook in hoofdstuk 3 een rol speelt, zullen wij deze apart vermelden.

VOORWAARDE (I): De familie kansmaten $\{P_\zeta, \zeta \in T \subset \mathbb{R}\}$ op (X,A) wordt gedomineerd door een σ-eindige maat μ. De dichtheid van P_ζ ten opzichte van μ noteren wij met g_ζ. Voorts is θ_0 een ophopingspunt van T en noteren wij P_{θ_0} en g_{θ_0} kort als P_0 en g_0. Er is een verzameling $X_0 \in A$, met $\mu(X_0^c) = 0$, zodat $g_0(x) > 0$ en $\dot{g}_0(x) := \partial g_\zeta(x)/\partial\zeta\big|_{\zeta=\theta_0}$ bestaat voor alle $x \in X_0$. Tenslotte geldt

$$(2.18) \qquad \int \dot{g}_0 d\mu = 0, \quad 0 < \int (\dot{g}_0^2/g_0) d\mu = J_0 < \infty,$$

$$(2.19) \qquad \lim_{\zeta\to\theta_0} \int (g_\zeta^{\frac{1}{2}} - g_0^{\frac{1}{2}})^2 (\zeta-\theta_0)^{-2} d\mu =$$

$$= \int \lim_{\zeta\to\theta_0} (g_\zeta^{\frac{1}{2}} - g_0^{\frac{1}{2}})^2 (\zeta-\theta_0)^{-2} d\mu \qquad (= J_0/4).$$

De grootheid J_0, die in (2.18) optreedt, noemt men wel de *Fisher-informatie* in θ_0 van de familie dichtheden $\{g_\zeta, \zeta \in T\}$.

STELLING 2.5: Laat de familie $\{P_\zeta, \zeta \in T \subset \mathbb{R}\}$ aan voorwaarde (I) voldoen. Bovendien is gegeven dat θ_0 en $\tau(N,n)$ voldoen aan

$$(2.20) \qquad \max_{1\le n\le N} [\tau(N,n) - \theta_0]^2 \to 0, \ \sum_{n=1}^{N} [\tau(N,n) - \theta_0]^2 \to \eta^2,$$

voor zekere $0 < \eta < \infty$, als $N \to \infty$. Zij P_0^N de productmaat $\prod_{n=1}^{N} P_0$ en zij

$$S_{Nn} := [\tau(N,n) - \theta_0][\dot{g}_0(X_{Nn})/g_0(X_{Nn})],$$

$$S_N := \sum_{n=1}^{N} S_{Nn},$$

voor $N = 1,2,\ldots$ en $n = 1,\ldots,N$.

Dan geldt $\{P_0^N\} \lhd\rhd \{P_{\tau(N)}\}$ en bovendien, als $N \to \infty$,

$$(2.21) \qquad L(S_N|P_0^N) \longrightarrow_z N(0;\eta^2 J_0),$$

$$(2.22) \qquad S_N - \Lambda_N \longrightarrow_{P_0^N} \tfrac{1}{2}\eta^2 J_0,$$

$$(2.23) \qquad L(\Lambda_N|P_0^N) \longrightarrow_z N(-\tfrac{1}{2}\eta^2 J_0;\eta^2 J_0).$$

Kies $\{c_N\}$ zodanig dat $c_N \to \eta J_0^{\frac{1}{2}}\xi_{1-\alpha}$, als $N \to \infty$, voor zekere $0 < \alpha < 1$. Dan is de rij toetsen $\{\phi_N\} = \{\chi(\{S_N > c_N\})\}$ asymptotisch meest onderscheidend voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$, met $\theta(N,n) = \theta_0$ voor $N = 1,2,\ldots$ en $n = 1,\ldots,N$. Het asymptotisch onderscheidingsvermogen is gelijk aan

$$(2.24) \qquad \lim_{N\to\infty} \int \phi_N dP_{\tau(N)} = 1 - \Phi(\xi_{1-\alpha} - \eta J_0^{\frac{1}{2}}).$$

BEWIJS:

Daar $\Lambda_N = S_N + (\Lambda_N - S_N)$, volgt (2.23) direct uit (2.21) en (2.22). Zij $\sigma^2 := \eta^2 J_0$.

Wij bewijzen eerst (2.21). Onder P_0^N is S_N een som van N onderling onafhankelijke stochastische grootheden S_{Nn}, die wegens de eerste relatie in (2.18) voldoen aan $E_{\theta(N)}S_{Nn} = 0$ en $Var_{\theta(N)}S_{Nn} = [\tau(N,n) - \theta_0]^2 J_0$, zodat uit de tweede relatie in (2.20) volgt

$$\lim_{N\to\infty} Var_{\theta(N)} S_N = \sigma^2.$$

Als $\sigma^2 = 0$, dan volgt uit de ongelijkheid van Chebyshev dat $S_N \longrightarrow_{P_0^N} 0$, zodat (2.21) dan juist is. Onderstel nu dat $\sigma^2 > 0$. Definieer

$$\tilde{S}_{Nn} := \dot{g}_0(X_{Nn})/g_0(X_{Nn}),$$

en kies een vaste $\varepsilon > 0$. Dan is voor $N \to \infty$

$$[\sigma^2 + o(1)]^{-1} \sum_{n=1}^{N} E_{\theta(N)} S_{Nn}^2 [\chi(\{t \geq \varepsilon\sigma + o(1)\})(|S_{Nn}|)] =$$

$$= [\sigma^2 + o(1)]^{-1} \sum_{n=1}^{N} [\tau(N,n) - \theta_0]^2 \times$$

$$\times E_{\theta(N)} \tilde{S}_{Nn}^2 [\chi(\{[\tau(N,n) - \theta_0]t \geq \varepsilon\sigma + o(1)\})(|\tilde{S}_{Nn}|)] \to 0.$$

Immers enerzijds geldt $\sum_{n=1}^{N} [\tau(N,n) - \theta_0]^2 \to \eta^2 < \infty$ wegens (2.20) en anderzijds volgt, eveneens uit (2.20),

$$\cup_{n=1}^{N} \{[\tau(N,n) - \theta_0]t \geq \varepsilon\sigma + o(1)\} \to \emptyset,$$

zodat volgens de gedomineerde convergentiestelling (met majorant $(\tilde{S}_{Nn})^2$, zie (2.18))

$$\max_{1 \leq n \leq N} E_{\theta(N)} \tilde{S}_{Nn}^2 [\chi(\{[\tau(N,n) - \theta_0]t \geq \varepsilon\sigma + o(1)\})(|\tilde{S}_{Nn}|)] \to 0.$$

Aan de Lindeberg-voorwaarde is derhalve voldaan, zodat (2.21) geldt op grond van de centrale limietstelling van Lindeberg-Feller.

Vervolgens leiden wij (2.22) af uit (2.21) en stelling 2.3. Laten W_{Nn} en W_N gedefinieerd zijn als in stelling 2.3. Dan geldt

$$\lim_{N \to \infty} E_{\theta(N)} W_N = -\sigma^2/4. \tag{2.25}$$

Immers, wegens (2.19) en (2.20) is

$$E_{\theta(N)} W_N = \sum_{n=1}^{N} E_{\theta(N)} W_{Nn} = 2 \sum_{n=1}^{N} \int [(g_{\tau(N,n)}/g_0)^{\frac{1}{2}} - 1] g_0 d\mu =$$

$$= -\sum_{n=1}^{N} \int [2g_0 - 2g_{\tau(N,n)}^{\frac{1}{2}} g_0^{\frac{1}{2}}] d\mu =$$

$$= -\sum_{n=1}^{N} \int [g_{\tau(N,n)}^{\frac{1}{2}} - g_0^{\frac{1}{2}}]^2 d\mu =$$

$$= -\sum_{n=1}^{N} [\tau(N,n)-\theta_0]^2 \int [g_{\tau(N,n)}^{\frac{1}{2}} - g_0^{\frac{1}{2}}]^2 [\tau(N,n)-\theta_0]^{-2} d\mu \to$$

$$\to -\eta^2 J_0/4.$$

Voorts is

$$\lim_{N\to\infty} E_{\theta(N)}(W_N - E_{\theta(N)}W_N - S_N)^2 = 0, \tag{2.26}$$

omdat voor $N \to \infty$

$$E_{\theta(N)}(W_N - E_{\theta(N)}W_N - S_N)^2 = Var_{\theta(N)}(W_N - S_N) =$$

$$= \sum_{n=1}^{N} Var_{\theta(N)}(W_{Nn} - S_{Nn}) \leq \sum_{n=1}^{N} E_{\theta(N)}(W_{Nn} - S_{Nn})^2 =$$

$$= 4 \sum_{n=1}^{N} E_{\theta(N)}\{[g_{\tau(N,n)}(X_{Nn})/g_0(X_{Nn})]^{\frac{1}{2}} - 1 - \tfrac{1}{2}[\tau(N,n)-\theta_0]\tilde{S}_{Nn}\}^2 =$$

$$= 4 \sum_{n=1}^{N} [\tau(N,n)-\theta_0]^2 E_{\theta(N)}\{[(g_{\tau(N,n)}(X_{Nn}))^{\frac{1}{2}} - (g_0(X_{Nn}))^{\frac{1}{2}}] \times$$

$$\times [\tau(N,n)-\theta_0]^{-1}[g_0(X_{Nn})]^{-\frac{1}{2}} - \tfrac{1}{2}\dot{g}_0(X_{Nn})/g_0(X_{Nn})\}^2 \to 0,$$

waarbij de limietovergang met behulp van lemma 2.4 volgt uit (2.19), (2.20) en de convergentie b.o. naar 0 van de vorm tussen accolades in het laatste lid.

Uit (2.26) volgt dat voor $N \to \infty$

$$W_N - E_{\theta(N)}W_N - S_N \xrightarrow[P_0^N]{} 0. \tag{2.27}$$

Daar $W_N = (W_N - E_{\theta(N)}W_N - S_N) + E_{\theta(N)}W_N + S_N$, impliceert deze relatie samen met (2.21) en (2.25) dat

$$L(W_N|P_0^N) \longrightarrow_z N(-\sigma^2/4;\sigma^2).$$

Bovendien is aan voorwaarde (2.11) van stelling 2.3 voldaan. Immers, voor elke $0 < \varepsilon < 1$ is

$$\{|g_{\tau(N,n)}/g_0 - 1| > \varepsilon\} \subset \{|(g_{\tau(N,n)}/g_0)^{\frac{1}{2}} - 1| > \varepsilon/3\},$$

zodat op grond van de ongelijkheid van Chebyshev, (2.19) en (2.20) volgt

$$\lim_{N\to\infty} \max_{1\le n\le N} P_0(|g_{\tau(N,n)}(X_{Nn})/g_0(X_{Nn})-1|\ge\varepsilon) \le$$

$$\le \limsup_{N\to\infty} \max_{1\le n\le N} P_0(|(g_{\tau(N,n)}(X_{Nn})/g_0(X_{Nn}))^{\frac{1}{2}}-1|\ge\varepsilon/3) \le$$

$$\le \limsup_{N\to\infty} \max_{1\le n\le N} 9\varepsilon^{-2} E_{\theta(N)}[(g_{\tau(N,n)}(X_{Nn})/g_0(X_{Nn}))^{\frac{1}{2}} - 1]^2 =$$

$$= 9\varepsilon^{-2} \limsup_{N\to\infty} \max_{1\le n\le N}[\tau(N,n) - \theta_0]^2 \times$$

$$\times \int [g^{\frac{1}{2}}_{\tau(N,n)} - g_0^{\frac{1}{2}}]^2[\tau(N,n) - \theta_0]^{-2}d\mu \to 0.$$

Derhalve volgt uit stelling 2.3 dat

$$W_N - \Lambda_N \longrightarrow_{P_0^N} \sigma^2/4.$$

Maar omdat $S_N - \Lambda_N = (S_N - W_N + E_{\theta(N)}W_N) + (W_N - \Lambda_N) - E_{\theta(N)}W_N$, leiden wij uit (2.25), (2.27) en bovenstaande relatie direct af dat

$$S_N - \Lambda_N \longrightarrow_{P_0^N} \tfrac{1}{2}\sigma^2.$$

De relaties (2.21) en (2.22) impliceren onmiddellijk (2.23) en daarmee $\{P_0^N\} \Leftrightarrow \{P_{\tau(N)}\}$.

De laatste beweringen van de stelling betreffende de asymptotische optimaliteit en het asymptotische onderscheidingsvermogen van de rij toetsen $\{\phi_N\}$ volgen onmiddellijk uit (2.22), (2.23) en stelling 2.2. □

In het volgende zullen wij voorwaarde (2.20) wat nader toelichten. In voorbeeld 2.1 zagen wij reeds dat de afstand tussen de parameters, corresponderende met het alternatief, en de parameter, corresponderende met de nulhypothese, niet te snel maar ook niet te langzaam naar 0 dient te convergeren opdat het asymptotisch onderscheidingsvermogen strikt tussen α en 1 ligt. Deze eis herkennen wij voor een wat algemenere situatie in (2.20). Wij zullen nu echter laten zien dat de juiste orde van de convergentie-snelheid van de parameters in wezen bepaald wordt door het gedrag van de Hellinger-afstand (zie definitie 1.10) tussen de kansmaten, behorende bij deze parameters. In het geval van stelling 2.5 kunnen, onder voorwaarde (I), de eisen opgelegd aan de parameters herleid worden tot analoge voorwaarden voor de Hellinger-afstand.

Zij voorwaarde (I) vervuld. Dan is de functie

$$\Delta(\zeta) := \delta(P_{\theta_0}, P_\zeta) \tag{2.28}$$

differentieerbaar in θ_0 op grond van (2.19). Zij $\dot{\Delta}_0 := \Delta'(\theta_0)$, dan is wegens (2.18)

$$0 < |\dot{\Delta}_0| < \infty. \tag{2.29}$$

LEMMA 2.6: Zij voorwaarde (I) vervuld. Dan is (zie (2.20))

$$\max_{1\le n\le N} [\tau(N,n)-\theta_0]^2 \to 0, \quad \sum_{n=1}^{N} [\tau(N,n)-\theta_0]^2 \to \eta^2$$

(voor $N\to\infty$) equivalent met

$$\max_{1\le n\le N} [\delta(P_{\theta_0}, P_{\tau(N,n)})]^2 \to 0, \tag{2.30}$$

$$\sum_{n=1}^{N} [\delta(P_{\theta_0}, P_{\tau(N,n)})]^2 \to \dot{\Delta}_0^2 \cdot \eta^2.$$

BEWIJS:
Zij (2.20) vervuld. Dan is

$$[\Delta(\zeta)]^2 = (\zeta-\theta_0)^2 \dot{\Delta}_0^2 + o((\zeta-\theta_0)^2),$$

als $\zeta \to \theta_0$, aangezien $\Delta(\theta_0) = 0$. Laten we nu voor ζ de waarden $\tau(N,n)$ substitueren. Door het maximum over n te nemen vinden wij

$$\max_{n=1,\ldots,N} [\Delta(\tau(N,n))]^2 = \dot{\Delta}_0^2 \max_{n=1,\ldots,N} [\tau(N,n)-\theta_0]^2 + o(1),$$

als $N\to\infty$. Sommeren over n geeft

$$\sum_{n=1}^{N}[\Delta(\tau(N,n))]^2 = \dot{\Delta}_0^2 \sum_{n=1}^{N}[\tau(N,n)-\theta_0]^2 + o(1),$$

als $N\to\infty$. Uit deze beide relaties volgt onmiddellijk (2.30).

Zij nu (2.30) gegeven. Zij $T(\cdot)$ de inverse functie van $\Delta(\cdot)$. Op grond van (2.29) bestaat $T(\Delta)$ in een omgeving van $\Delta = \Delta(\theta_0) = 0$ en is $T'(0) = 1/\dot{\Delta}_0$. Hieruit volgt dat voor $\Delta \to 0$

$$[T(\Delta)-T(0)]^2 = \Delta^2 \cdot (1/\dot{\Delta}_0)^2 + o(\Delta^2).$$

Door substitutie van $\Delta = \Delta(\tau(N,n))$ in deze relatie leidt men thans (2.20) op dezelfde wijze als boven af uit (2.30). □

In (1.12) hebben we reeds gezien dat $[\delta(P,Q)]^2 = \int(p^{\frac{1}{2}}-q^{\frac{1}{2}})d\mu =$ $= 2 - 2\int(pq)^{\frac{1}{2}}d\mu$. Uit deze relatie volgt nu onmiddellijk het volgende verband tussen de Hellinger-afstand van productmaten en de Hellinger-afstand van hun componenten

$$(2.31) \qquad [\delta(P_{\theta(N)},P_{\tau(N)})]^2 = 2 - 2\prod_{n=1}^{N}(1-\tfrac{1}{2}[\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2).$$

Uit lemma 2.6 blijkt dat voorwaarde (2.20) in stelling 2.5 vervangen kan worden door (2.30). In feite is (2.30) een bijzonder geval van de iets algemenere voorwaarde

$$(2.32) \qquad \max_{n=1,\ldots,N} [\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2 \to 0,$$

$$\sum_{n=1}^{N} [\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2 \to a^2,$$

voor zekere constante $0 < a < \infty$, als $N\to\infty$. We zullen nu laten zien dat deze voorwaarde impliceert dat

(2.33) $$0 < \lim_{N\to\infty} \delta(P_{\theta(N)},P_{\tau(N)}) = [2-2\exp(-\tfrac{1}{2}a^2)]^{\frac{1}{2}} < 2^{\frac{1}{2}}.$$

Het is voldoende om aan te tonen dat

(2.34) $$\lim_{N\to\infty} \sum_{n=1}^{N} \log(1-\tfrac{1}{2}[\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2) = -\tfrac{1}{2}a^2,$$

op grond van (2.31). Uit (2.32) volgt dat de logarithme in (2.34) gedefinieerd is voor voldoend grote N. Nu is, voor $N\to\infty$,

$$\sum_{n=1}^{N} \log(1-\tfrac{1}{2}[\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2) =$$

$$= -\tfrac{1}{2}\sum_{n=1}^{N} [\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2 + o(\sum_{n=1}^{N}[\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^4).$$

Het gegeven impliceert

$$\sum_{n=1}^{N} [\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^4 \leq$$

$$\leq \max_{n=1,\ldots,N} [\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2 \times$$

$$\times \sum_{n=1}^{N} [\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2 \to 0,$$

als $N\to\infty$, waaruit (2.34) onmiddellijk volgt.

Voorwaarde (2.32) sluit dus kennelijk uit dat $\lim_{N\to\infty} \delta(P_{\theta(N)},P_{\tau(N)} = 2^{\frac{1}{2}}$. Uit de volgende stelling blijkt dat deze laatste situatie ook niet erg interessant is.

STELLING 2.7: Stel dat $\lim_{N\to\infty} \delta(P_{\theta(N)},P_{\tau(N)}) = 2^{\frac{1}{2}}$. Dan bestaan er voor iedere $0 \leq \alpha \leq 1$ consistente rijen toetsen voor de problemen $(\alpha;\{\theta(N)\},\{\tau(N)\})$ en $(\alpha;\{\tau(N)\},\{\theta(N)\})$.

BEWIJS:
Het is voldoende de stelling te bewijzen voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ met $\alpha = 0$. Laten $0 < a \leq b < \infty$ twee willekeurige constanten zijn en zij $\{k_N\}$ een rij getallen zodat $a \leq k_N \leq b$ voor alle N. Definieer de rij toetsen $\{\phi_N\}$ door

$$\phi_N(x_1,\ldots,x_N) = \begin{cases} 0 & \text{als } \Pi_{n=1}^N [g_{\tau(N,n)}(x_n)]^{\frac{1}{2}} \leq k_N \Pi_{n=1}^N [g_{\theta(N,n)}(x_n)]^{\frac{1}{2}} \\ 1 & \text{als } \Pi_{n=1}^N [g_{\tau(N,n)}(x_n)]^{\frac{1}{2}} > k_N \Pi_{n=1}^N [g_{\theta(N,n)}(x_n)]^{\frac{1}{2}} \end{cases}$$

Uit het gegeven en (2.29) volgt dat

$$\int \phi_N dP_{\theta(N)} = \int \phi_N(x_1,\ldots,x_N) \Pi_{n=1}^N g_{\theta(N,n)}(x_n) d\mu_N(x_1,\ldots,x_N) \leq$$

$$\leq (k_N)^{-1} \int [\Pi_{n=1}^N g_{\theta(N,n)}(x_n) g_{\tau(N,n)}(x_n)]^{\frac{1}{2}} d\mu_N(x_1,\ldots,x_N) =$$

$$= (k_N)^{-1}(1-\tfrac{1}{2}[\delta(P_{\theta(N)},P_{\tau(N)})]^2) \to 0,$$

als $N \to \infty$. Analoog bewijst men dat $\int (1-\phi_N) dP_{\tau(N)} \to 0$, als $N \to \infty$. □

Aan de voorwaarden van stelling 2.7 is zeker voldaan als $\theta(N,1) = \ldots = \theta(N,N) = \theta_0$ en $\tau(N,1) = \ldots = \tau(N,N) = \theta_1$, met $\theta_0 \neq \theta_1$.

Uit stelling 2.7 zien wij dat de voorwaarde $\delta(P_{\theta(N)},P_{\tau(N)}) \not\to 2^{\frac{1}{2}}$ nodig is opdat b.v. $\{P_{\tau(N)}\} \lhd \{P_{\theta(N)}\}$. Immers, als $\{\phi_N\}$ de rij toetsen uit het bewijs van stelling 2.7 is, geldt $P_{\theta(N)}(\{\phi_N>0\}) \to 0$ terwijl $P_{\tau(N)}(\{\phi_N>0\}) \to 1$. Een nodige en voldoende voorwaarde voor eenzijdige naburigheid verkrijgen wij indien wij eisen, dat iedere rij toetsen met asymptotische onbetrouwbaarheid 0 ook asymptotisch onderscheidingsvermogen 0 heeft.

STELLING 2.8: De volgende twee uitspraken zijn equivalent:

(a) voor iedere rij toetsen $\{\psi_N\}$ voor het probleem $(0;\{\theta(N)\},\{\tau(N)\})$ geldt dat $\int \psi_X dP_{\tau(N)} \to 0$, als $N \to \infty$;

(b) $\{P_{\tau(N)}\} \lhd \{P_{\theta(N)}\}$.

BEWIJS:

Stel dat aan (a) is voldaan en dat $P_{\theta(N)}(A_N) \to 0$, als $N \to \infty$ (voor $A_N \in A^N$). Kies $\{\psi_N\} = \{\chi(A_N)\}$. Aangezien dit een rij toetsen voor het probleem

$(0;\{\theta(N)\},\{\tau(N)\})$ is, volgt uit het gegeven dat $\int\chi(A_N)dP_{\tau(N)} = P_{\tau(N)}(A_N) \to 0$, als $N \to \infty$.

Laat nu omgekeerd aan (b) voldaan zijn en stel dat $\{\psi_N\}$ een rij toetsen is voor het probleem $(0;\{\theta(N)\},\{\tau(N)\})$. Hieruit volgt dat $P_{\theta(N)}(\{\psi_N \geq \varepsilon\}) \to 0$, als $N \to \infty$, voor iedere $\varepsilon > 0$. Op grond van het gegeven houdt de laatste bewering in, dat $P_{\tau(N)}(\{\psi_N \geq \varepsilon\}) \to 0$, als $N \to \infty$, voor iedere $\varepsilon > 0$. Hieruit volgt dat $\int\psi_N dP_{\tau(N)} \to 0$, als $N \to \infty$. □

De karakterisering van eenzijdige naburigheid in stelling 2.8 berust op rijen toetsen met asymptotische onbetrouwbaarheid 0. Hoewel in asymptotische problemen de voorwaarde, dat de asymptotische onbetrouwbaarheid 0 is, zeker redelijk is, hebben wij ons toch tot asymptotische onbetrouwbaarheidsdrempels $0 < \alpha < 1$ beperkt.

GEVOLG 2.9: De volgende implicatie geldt:

$$\lim_{N\to\infty} \delta(P_{\theta(N)},P_{\tau(N)}) \to 2^{\frac{1}{2}} \Longrightarrow \begin{cases} \{P_{\theta(N)}\} \not\triangleleft \{P_{\tau(N)}\}, \\ \\ \{P_{\tau(N)}\} \not\triangleleft \{P_{\theta(N)}\}. \end{cases}$$

BEWIJS:
Dit volgt onmiddellijk uit stelling 2.7 en 2.8. □

Tot besluit van dit hoofdstuk zullen wij het begrip asymptotische relatieve doeltreffendheid van een rij toetsen met betrekking tot een tweede rij toetsen invoeren. Wij geven alleen een definitie voor het geval, dat beide rijen toetsen een asymptotisch onderscheidingsvermogen van de vorm (2.8) hebben.

DEFINITIE 2.3: Voor $i = 1,2$ zij $\{\phi_{iN}\}$ een rij toetsen voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$, waarin $0 < \alpha < 1$, met asymptotisch onderscheidingsvermogen $1 - \Phi(\xi_{1-\alpha}-\sigma_i)$ voor zekere $\sigma_i > 0$. De grootheid

$$(2.35) \qquad ARE(\alpha;\{\phi_{1N}\},\{\phi_{2N}\};\{\tau(N)\}) := (\sigma_1/\sigma_2)^2$$

noemt men de *asymptotische relatieve doeltreffendheid* (Engels: *asymptotic relative efficiency*, afgekort tot *ARE*) van de rij $\{\phi_{1N}\}$ met betrekking tot de rij $\{\phi_{2N}\}$ voor de asymptotische onbetrouwbaarheidsdrempel α en de rij alternatieven $\{\tau(N)\}$.

In de laatste stelling van dit hoofdstuk zullen wij nader ingaan op de asymptotische relatieve doeltreffendheid van een willekeurige rij toetsen met betrekking tot een asymptotisch meest onderscheidende rij toetsen.

STELLING 2.10: Laat T_{iN} een stochastische grootheid zijn op (X^N,A^N), voor $i = 1,2$ en alle N, zodat

$$(2.36) \qquad L((T_{1N},T_{2N}) \mid P_{\theta(N)}) \to_z N(0,0;1,1,\rho),$$

voor zekere $0 < \rho \leq 1$, terwijl bovendien

$$(2.37) \qquad \sigma T_{1N} - \Lambda_N \longrightarrow_{P_{\theta(N)}} \tfrac{1}{2}\sigma^2,$$

als $N \to \infty$, voor zekere $0 < \sigma < \infty$. Kies de rij $\{c_{iN}\}$ zodanig dat $c_{iN} \to \xi_{1-\alpha}$, als $N \to \infty$, voor zekere $0 < \alpha < 1$ en beschouw de rij toetsen $\{\phi_{iN}\} = \{\chi(\{T_{iN} > c_{iN}\})\}$, $i = 1,2$.

Dan is $\{\phi_{1N}\}$ een asymptotisch meest onderscheidende rij toetsen voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ en

$$\mathrm{ARE}(\alpha;\{\phi_{2N}\},\{\phi_{1N}\};\{\tau(N)\}) = \rho^2.$$

BEWIJS:
Uit (2.36) en (2.37) volgt, dat

$$L(\Lambda_N|P_{\theta(N)}) \longrightarrow_z N(-\tfrac{1}{2}\sigma^2;\sigma^2),$$

zodat alle conclusies van stelling 2.2 van toepassing zijn op de rij toetsen $\{\phi_{1N}\}$, die derhalve asymptotisch meest onderscheidend is en waarvoor

(2.38) $\lim_{N\to\infty} \int \phi_{1N} dP_{\tau(N)} = 1 - \Phi(\xi_{1-\alpha} - \sigma).$

Uit (2.36) en (2.37) volgt bovendien, dat

(2.39) $L((\Lambda_N, T_{2N}) \mid P_{\theta(N)}) \to_z N(-\tfrac{1}{2}\sigma^2, 0; \sigma^2, 1, \rho).$

Met behulp van stelling 1.19 kunnen wij uit (2.39) afleiden, dat

(2.40) $L(T_{2N} | P_{\tau(N)}) \longrightarrow_z N(\rho\sigma; 1).$

Uit het gegeven $L(T_{2N}|P_{\theta(N)}) \longrightarrow_z N(0;1)$ kunnen wij concluderen dat de rij toetsen $\{\phi_{2N}\}$ een asymptotische onbetrouwbaarheid gelijk aan α heeft en, in combinatie met (2.40), dat

(2.41) $\lim_{N\to\infty} \int \phi_{2N} dP_{\tau(N)} = 1 - \Phi(\xi_{1-\alpha} - \rho\sigma).$

De bewering betreffende de asymptotische relatieve doeltreffendheid volgt nu onmiddellijk uit (2.38), (2.41) en definitie 2.3. □

BRONVERMELDINGEN

De kern van dit hoofdstuk - stelling 2.1, 2.2, en 2.10 en in het bijzonder stelling 2.5 - is ontleend aan WITTING & NÖLLE [24]. De vorm waarin hier het zgn. tweede lemma van LeCam wordt gepresenteerd is iets algemener dan in [24]. De lezer wordt hiervoor naar HÁJEK & ŠIDÁK [12] verwezen. Als algemene referenties voor de asymptotische toetsingstheorie vermelden wij verder LECAM [16,17], HÁJEK [10] en ROUSSAS [22], ook reeds in de bronvermelding van het eerste hoofdstuk genoemd.

De relatie (2.31) tussen Hellinger-afstanden van productmaten en Hellinger-afstanden van hun marginalen kan men bijv. ook bij LECAM [18] vinden. LECAM [19] definieert de afstand tussen parameters als de Hellinger-afstand tussen de bijbehorende kansmaten. Onder voorwaarde (I) is dit afstandsbegrip equivalent met de Euclidische afstand tussen de parameters (zie lemma 2.6). Stelling 2.7 vormt in wezen een gedeelte van de inhoud van een door KAKUTANI [14] bewezen stelling; zie ook KRAFT [15]. Het resultaat van stelling 2.8 wordt vermeld door HÁJEK & ŠIDÁK [12] en ROUSSAS [22].

Hoofdstuk 3 ASYMPTOTISCHE OPTIMALITEIT VAN RANGTOETSEN

In het vorige hoofdstuk hebben we in stelling 2.5 gezien dat voor zekere asymptotische toetsingsproblemen $(\alpha;\{\theta(N)\},\{\tau(N)\})$, waarbij voor iedere $N = 1,2,\ldots$ een enkelvoudige nulhypothese $\theta(N)$ tegen een enkelvoudige alternatieve hypothese $\tau(N)$ getoetst wordt met een zodanige onbetrouwbaarheidsdrempel α_N dat $\limsup_{N\to\infty} \alpha_N \leq \alpha$, onder voorwaarden welke naburigheid van de onderliggende kansmaten impliceren, asymptotisch meest onderscheidende toetsen geconstrueerd kunnen worden. Omdat de met deze toetsen corresponderende toetsingsgrootheden afhankelijk zijn van de parameters $\theta(N)$ en $\tau(N)$, zijn deze toetsen in het algemeen niet bruikbaar voor het toetsen van een samengestelde nulhypothese tegen een samengestelde alternatieve hypothese.

In dit hoofdstuk zullen we laten zien dat "asymptotisch uniform meest onderscheidende" rijen toetsen voor enige bekende toetsingsproblemen uit de verdelingsvrije statistiek, waarbij een samengestelde nulhypothese tegen een samengestelde alternatieve hypothese getoetst wordt, gevonden kunnen worden onder de rangtoetsen.

Hieraan voorafgaand geven we een uiteenzetting over de (eindige) theorie der rangtoetsen. Meer in het bijzonder zullen we ingaan op de constructie van "lokaal meest onderscheidende" rangtoetsen.(zie bijvoorbeeld HÁJEK & ŠIDÁK [12]en WITTING & NÖLLE [24]).

De theorie over asymptotische optimaliteit van rangtoetsen werd grotendeels ontwikkeld door HÁJEK [10]. Uitvoerige beschouwingen over dit onderwerp zijn te vinden in HÁJEK & ŠIDÁK [12] en WITTING & NÖLLE [24].

We beginnen de theorie van dit hoofdstuk met de introductie van enige voor de theorie der rangtoetsen belangrijke begrippen en resultaten. Omdat we voorlopig uitsluitend te maken zullen hebben met de situatie dat het aantal waarnemingen N een vast doch willekeurig natuurlijk getal is, zullen we de index N in onze notatie meestal weglaten. Zo schrijven we $X = (X_1,\ldots,X_N)$ in plaats van $X_N = (X_{N1},\ldots,X_{NN})$ en $\zeta = (\zeta_1,\ldots,\zeta_N)$ en $P_\zeta = \Pi_{n=1}^N P_{\zeta_n}$ waar we $\zeta(N) = (\zeta(N,1),\ldots,\zeta(N,N))$ en $P_{\zeta(N)} = \Pi_{n=1}^N P_{\zeta(N,n)}$ bedoelen. Dikwijls schrijven we $P^N = \Pi_{n=1}^N P_n$ respectievelijk $Q^N = \Pi_{n=1}^N Q_n$

in plaats van $P_{\theta(N)} = \Pi_{n=1}^{N} P_{\theta(N,n)}$ respectievelijk $P_{\tau(N)} = \Pi_{n=1}^{N} P_{\tau(N,n)}$. Daar waar elementen $\theta(N) = (\theta(N,1),\ldots,\theta(N,N))$ en $\tau(N) = (\tau(N,1),\ldots,\tau(N,N))$ ter sprake komen schrijven we $\theta = (\theta_1,\ldots,\theta_N)$ respectievelijk $\tau = (\tau_1,\ldots,\tau_N)$.

Laten $X_1,\ldots,X_N$ o.o. stochastische grootheden zijn, met verdelingsfuncties $F_1,\ldots,F_N$ en bijbehorende kansverdelingen $P_1,\ldots,P_N$ op $(X,A) = (\mathbb{R}^k,\mathcal{B}^k)$, voor een zeker vast natuurlijk getal k. Voorlopig zullen we veronderstellen dat k = 1.

Bij een stochastische vector $X = (X_1,\ldots,X_N)$ met verdelingsfunctie $F^N = \Pi_{n=1}^{N} F_n$ en productmaat $P^N = \Pi_{n=1}^{N} P_n$ op $(X^N,A^N) = (\mathbb{R}^N,\mathcal{B}^N)$, en met waarden $x = (x_1,\ldots,x_N)$, definiëren we

$$X^{(\cdot)} = (X^{(1)},\ldots,X^{(N)})$$

als de bijbehorende stochastische vector der naar opklimmende grootte geordende componenten $X^{(1)},\ldots,X^{(N)}$, met waarden $x^{(1)},\ldots,x^{(N)}$; m.a.w.

$$x^{(1)} \leq x^{(2)} \leq \ldots \leq x^{(N)}.$$

Zij verder, voor vectoren $(x_1,\ldots,x_N)$, waarvan alle componenten verschillend zijn, r_n het rangnummer van x_n in de geordende rij $x^{(1)} < x^{(2)} < \ldots < x^{(N)}$;

$$x_n = x^{(r_n)}, \qquad n = 1,\ldots,N.$$

De stochastische grootheid $R_n = r_n(X_1,\ldots,X_N)$ heet het *rangnummer* van X_n (n=1,...,N). In het bijzonder belangrijke speciale geval dat $X_1,\ldots,X_N$ o.o. stochastische grootheden zijn en hun verdelingsfuncties $F_1,\ldots,F_N$ allen continu zijn, kunnen de waarden van $X_1,\ldots,X_N$ met kans 1 ondubbelzinnig naar grootte gerangschikt worden. De rangnummers $R_1,\ldots,R_N$ zijn dan met kans 1 goed gedefinieerd. De stochastische vector $R = (R_1,\ldots,R_N)$, de vector der rangnummers, heeft als mogelijke waarden $r = (r_1,\ldots,r_N)$ de N! permutaties van (1,2,...,N). De collectie van alle N! permutaties van (1,2,...,N) zullen we met $\mathcal{R}$ aanduiden. We merken nog op dat $X_1,\ldots,X_N$ uit $X^{(1)},\ldots,X^{(N)}$ en $R_1,\ldots,R_N$ gereconstrueerd kunnen worden. Er is een 1-1-

duidige correspondentie tussen $(x_1,\ldots,x_N)$ en $(x^{(1)},\ldots,x^{(N)})$ voor iedere waarde R = r van $\mathcal{R}$.

In het vervolg zullen we veelvuldig te maken krijgen met de situatie dat $X_1,\ldots,X_N$ o.o. identiek verdeelde stochastische grootheden zijn, ieder met continue verdelingsfunctie F en bijbehorende kansmaat P op $(\mathcal{X},\mathcal{A}) = (\mathbb{R},\mathcal{B})$. Dikwijls spreekt men dan van een *(aselecte) steekproef* van omvang N uit de kansverdeling P met verdelingsfunctie F. De bijbehorende stochastische vector $(X^{(1)},\ldots,X^{(N)})$ wordt dan de *geordende steekproef* (Engels: order statistics) genoemd.

LEMMA 3.1: (a) Laten $X_1,\ldots,X_N$ o.o. stochastische grootheden zijn, ieder met continue verdelingsfunctie F en kansverdeling P op $(\mathbb{R},\mathcal{B})$. Dan geldt

$$P^N(R=r) = P^N(\{(x_1,\ldots,x_N) : x_n=x^{(r_n)},\ n=1,\ldots,N\}) = \frac{1}{N!}$$

voor alle $r = (r_1,\ldots,r_N) \in \mathcal{R}$.

(b) $X^{(.)}$ en R zijn onafhankelijk.

(c) Indien bovendien P absoluut continu is t.o.v. een σ-eindige maat μ op $(\mathbb{R},\mathcal{B})$, met dichtheid f, dan wordt de simultane dichtheid van $X^{(1)},\ldots,X^{(N)}$, t.o.v. de productmaat μ_N op $(\mathbb{R}^N,\mathcal{B}^N)$, gegeven door

$$\begin{cases} N!\ \prod_{n=1}^{N} f(x^{(n)}) & \text{voor } x^{(1)} < x^{(2)} < \ldots < x^{(N)} \\ 0 & \text{anders.} \end{cases}$$

BEWIJS:

(a) Wegens de continuïteit van F kunnen we ons zonder bezwaar beperken tot punten $x \in \mathbb{R}^N$ met coördinaten $(x_1,\ldots,x_N)$ waarvoor $x^{(1)} < \ldots < x^{(N)}$. De deelverzameling {r} van $\mathcal{R}$ correspondeert juist met de verzameling van alle vectoren $(x_1,\ldots,x_N)$ met $x^{(1)} < \ldots < x^{(N)}$, waarvoor $x_n = x^{(r_n)}$, n = 1,...,N. Omdat de kansverdeling P^N van de stochastische vector $(X_1,\ldots,X_N)$ op grond van het gegeven invariant is m.b.t. permutaties van $X_1,\ldots,X_N$, geldt dat P^N (R=r) onafhankelijk is van

$r = (r_1,\ldots,r_N)$, de vector der rangnummers van $(x_1,\ldots,x_N)$. Daar de deelverzamelingen $\{R=r\}$ disjunct zijn voor verschillende waarden van $r \in \mathcal{R}$ en $\sum_{r\in\mathcal{R}} P^N(R=r) = 1$ volgt de bewering.

(b) Voor iedere Borelverzameling $B \in \mathcal{B}^N$ bestaande uit punten met coördinaten $x_1,\ldots,x_N$, waarvoor $x^{(1)} < \ldots < x^{(N)}$, geldt

$$(3.1) \qquad P^N((X^{(1)},\ldots,X^{(N)})\in B) =$$

$$= \sum_{r\in\mathcal{R}} P^N(R_1=r_1,\ldots,R_N=r_N \wedge (X^{(1)},\ldots,X^{(N)})\in B).$$

We merken verder op dat voor iedere $r = (r_1,\ldots,r_N) \in \mathcal{R}$ geldt

$$(3.2) \qquad P^N(R_1=r_1,\ldots,R_N=r_N \wedge (X^{(1)},\ldots,X^{(N)})\in B) =$$

$$= \int \chi(\{(x_1,\ldots,x_N) : x_n = x^{(r_n)}, n=1,\ldots,N, (x^{(1)},\ldots,x^{(N)})\in B\})\,dP^N =$$

$$= \int \chi(\{(x_1,\ldots,x_N) : x_n = x^{(n)}, n=1,\ldots,N, (x^{(1)},\ldots,x^{(N)})\in B\})\,dP^N =$$

$$= P^N(R_1=1,\ldots,R_N=N \wedge (X^{(1)},\ldots,X^{(N)})\in B)$$

waarbij de eerste en derde overgang triviaal zijn en de tweede overgang volgt uit het feit dat de kansverdeling P^N invariant is met betrekking tot de permutaties van $X_1,\ldots,X_N$.

Uit (3.1) en (3.2) volgt nu dat

$$P^N((X^{(1)},\ldots,X^{(N)})\in B) = N!P^N(R_1=r_1,\ldots,R_N=r_N \wedge (X^{(1)},\ldots,X^{(N)})\in B)$$

voor iedere $r = (r_1,\ldots,r_N) \in \mathcal{R}$, zodat wegens lemma 3.1.(a) volgt dat

$$P^N((X^{(1)},\ldots,X^{(N)})\in B).P^N(R=r) =$$

$$= P^N(R_1=r_1,\ldots,R_N=r_N \wedge (X^{(1)},\ldots,X^{(N)})\in B)$$

voor iedere $r = (r_1,\ldots,r_N) \in \mathcal{R}$ en iedere $B \in \mathcal{B}^N$ bestaande uit punten met coördinaten $(x_1,\ldots,x_N)$ waarvoor $x^{(1)} < \ldots < x^{(N)}$.

(c) Het bewijs van dit onderdeel van het lemma volgt onmiddellijk uit het feit dat $P^N((X^{(1)},\ldots,X^{(N)})\in B) = N!P^N((X_1,\ldots,X_N)\in B)$ geldt voor iedere $B \in \mathcal{B}^N$, bestaande uit punten met coördinaten $(x_1,\ldots,x_N)$ waarvoor $x_1 < \ldots < x_N$, wegens de invariantie van de kansverdeling P^N m.b.t. permutaties van $X_1,\ldots,X_N$. □

Lemma 3.1 zal gebruikt worden bij de berekening van de verdeling van een rangtoetsingsgrootheid onder de nulhypothese. Voor het berekenen van het onderscheidingsvermogen van een rangtoets is het volgende lemma van grote betekenis.

LEMMA 3.2: (a) Laten $X_1,\ldots,X_N$ o.o. stochastische grootheden zijn, ieder met continue verdelingsfunctie F en kansverdeling P op $(\mathbb{R},\mathcal{B})$. Zij verder $T(X_1,\ldots,X_N)$ een stochastische grootheid, dan geldt

$$E(T(X_1,\ldots,X_N)\,|\,R=r) = E(T(X^{(r_1)},\ldots,X^{(r_N)})). \tag{3.3}$$

(b) Indien $Q_1,Q_2,\ldots,Q_N$ kansverdelingen met verdelingsfuncties $G_1,G_2,\ldots,G_N$ op $(\mathbb{R},\mathcal{B})$ zijn, welke absoluut continu zijn t.o.v. P, en μ een σ-eindige maat op $(\mathbb{R},\mathcal{B})$ is die P en $Q_1,\ldots,Q_N$ domineert, dan geldt

$$Q^N(R=r) = \frac{1}{N!}\,E\prod_{n=1}^{N}\frac{g_n(X^{(r_n)})}{f(X^{(r_n)})} \tag{3.4}$$

waarbij $Q^N = \prod_{n=1}^{N} Q_n$ is, $g_1,\ldots,g_N$, f dichtheden van $Q_1,\ldots,Q_N$ respectievelijk P t.o.v. μ zijn en $X^{(1)},\ldots,X^{(N)}$ een geordende steekproef uit de kansverdeling P voorstelt.

BEWIJS:

(a) Daar $X_n = X^{(r_n)}$, $n=1,\ldots,N$, als $R = r$, en bovendien $X^{(.)}$ en R onafhankelijk zijn wegens lemma 3.1.(b), geldt

$$E(T(X_1,\ldots,X_N)\,|\,R=r) = E(T(X^{(r_1)},\ldots,X^{(r_N)})\,|\,R=r) = E(T(X^{(r_1)},\ldots,X^{(r_N)})).$$

(b) $Q^N(R=r) =$

$$= \int \chi(\{(x_1,\ldots,x_N) : x_n = x^{(r_n)}, n=1,\ldots,N\}) \prod_{n=1}^N g_n(x_n) d\mu_N(x_1,\ldots,x_N) =$$

$$= \frac{1}{N!} \int \chi(\{(x_1,\ldots,x_N) : x_n = x^{(r_n)}, n=1,\ldots,N\}) N! \prod_{n=1}^N \{\frac{g_n(x_n)}{f(x_n)} f(x_n)\} d\mu_N(x_1,\ldots,x_N),$$

waarbij wegens de absolute continuïteit van Q_n t.o.v. P het quotient $\frac{g_n}{f}$ P-b.o. de dichtheid van Q_n t.o.v. P voorstelt. Verder volgt uit lemma 3.1.(c) dat $N! \prod_{n=1}^N f(x^{(r_n)})$ de dichtheid van $X^{(r_1)},\ldots,X^{(r_N)}$ t.o.v. de productmaat μ_N voorstelt. Hiermee is het bewijs voltooid. □

OPMERKING 3.1: Formule (3.4) staat in de literatuur bekend als de formule van Hoeffding.

Na deze inleiding geven we nu de definitie van een rangtoets. Hoewel deze definitie bijzonder eenvoudig is, moet zij voor verschillende hypothesen afzonderlijk gegeven worden. We geven de definitie van een rangtoets voor drie belangrijke hypothesen uit de verdelingsvrije statistiek.

Laten, zoals reeds eerder is gezegd, $X_1,\ldots,X_N$ o.o. stochastische grootheden zijn, met verdelingsfuncties $F_1,\ldots,F_N$ en kansverdelingen $P_1,\ldots,P_N$ op $(X,A) = (\mathbb{R}^k,\mathcal{B}^k)$ voor een zeker vast natuurlijk getal k. Zij $\mathcal{F}_C$ de klasse van alle continue verdelingsfuncties op $\mathbb{R}$ en $\mathcal{F}_S$ de klasse van alle continue symmetrische verdelingsfuncties op $\mathbb{R}$ met symmetriepunt 0, d.w.z. $F \in \mathcal{F}_C$ en $F(x) = 1 - F(-x)$, $-\infty < x < \infty$.

De hypothese H_0 (isomorie) [1] Zij k = 1 en

$$H_0 = \{(F_1,\ldots,F_N) : F_1=\ldots=F_N=F\in\mathcal{F}_C\},$$

---

[1] In de Engelstalige literatuur spreekt men over de hypothese H_0 van "randomness".

dan heet een toets voor H_0 een rangtoets voor H_0, indien de bijbehorende kritieke functie ϕ een functie is, die slechts van de vector $R=(R_1,\ldots,R_N)$ der rangnummers afhangt.

De hypothese H_1 (symmetrie). Zij k = 1 en

$$H_1 = \{(F_1,\ldots,F_N) : F_1 = \ldots = F_N = F \in \mathcal{F}_S\}.$$

Definieer naast de stochastische grootheden $X_1,\ldots,X_N$ ook nog de stochastische grootheden sgn $X_1,\ldots,$sgn X_N, waarbij

$$\begin{cases} = 1 & \text{als } x > 0 \\ \operatorname{sgn} x = 0 & \text{als } x = 0 \\ = -1 & \text{als } x < 0, \end{cases}$$

de absolute waarden $|X_1|,\ldots,|X_N|$ en de geordende absolute waarden $|X|^{(1)},\ldots,|X|^{(N)}$ van $X_1,\ldots,X_N$. Verder definiëren we rangnummers R_n^+, n = 1,...,N van de absolute waarden door $R_n^+ = R_n(|X|)$, n = 1,...,N.

Dan heet een toets voor H_1 een rangtoets voor H_1, indien de bijbehorende kritieke functie ϕ een functie is, die slechts van de vectoren $R^+ = (R_1^+,\ldots,R_N^+)$ en sgn X = (sgn $X_1,\ldots,$sgn X_N) afhangt.

De hypothese H_2 (onafhankelijkheid). Zij k = 2 en

$$H_2 = \{(F_1,\ldots,F_N) : F_1 = \ldots = F_N = F,\ F = H.K,\ H,K \in \mathcal{F}_C\}.$$

$X_1 = (Y_1,Z_1),\ldots,X_N = (Y_N,Z_N)$ zijn o.o. stochastische vectoren, met continue tweedimensionale verdelingsfuncties $F_1,\ldots,F_N$ en kansverdelingen $P_1,\ldots,P_N$ op $(\mathbb{R}^2,\mathcal{B}^2)$. De stochastische vector $X = (X_1,\ldots,X_N)$ heeft dan verdelingsfunctie $\Pi_{n=1}^N F_n$ en kansverdeling $P^N = \Pi_{n=1}^N P_n$ op $(X^N,\mathcal{A}^N) = (\mathbb{R}^{2N},\mathcal{B}^{2N})$. Indien nu $R_1,\ldots,R_N$ de rangnummers van $Y_1,\ldots,Y_N$ en $S_1,\ldots,S_N$ die van $Z_1,\ldots,Z_N$ voorstellen, dan heet een toets voor H_2 een rangtoets voor H_2, indien de bijbehorende kritieke functie ϕ een functie is die slechts van de vectoren

$R = (R_1,\ldots,R_N)$ en $S = (S_1,\ldots,S_N)$ van rangnummers afhangt.

We merken nog op dat lemma 3.1 speciaal geschikt is voor het bestuderen van rangtoetsen voor de hypothese H_0. Analoge resultaten met het oog op rangtoetsen voor de hypothesen H_1 en H_2 zijn geformuleerd in de volgende twee lemma's.

LEMMA 3.3: (a) Laten $X_1,\ldots,X_N$ o.o. stochastische grootheden zijn, ieder met continue symmetrische verdelingsfunctie F met symmetriepunt 0 en kansverdeling P op $(\mathbb{R},\mathcal{B})$. Dan geldt

$$P^N(\text{sgn } X=v) = 2^{-N}$$

voor alle $v = (v_1,\ldots,v_N)$ met $v_n = 1$ of -1, $n = 1,\ldots,N$, en

$$P^N(R^+=r) = \frac{1}{N!}$$

voor alle $r = (r_1,\ldots,r_N) \in \mathcal{R}$.

(b) sgn X, R^+ en $|X|^{(.)}$ zijn onderling onafhankelijk.

(c) Indien bovendien P absoluut continu is t.o.v. een σ-eindige maat μ op $(\mathbb{R},\mathcal{B})$, met dichtheid f, dan wordt de simultane dichtheid van $|X|^{(1)},\ldots,|X|^{(N)}$, t.o.v. de productmaat μ_N op $(\mathbb{R}^N,\mathcal{B}^N)$, gegeven door

$$\begin{cases} 2^N\, N!\, \prod_{n=1}^N f(x^{(n)}) & \text{voor } 0 < x^{(1)} < \ldots < x^{(N)} \\ 0 & \text{anders.} \end{cases}$$

BEWIJS:
Het bewijs volgt direct uit lemma 3.1 en de symmetrie van de verdelingsfunctie F. □

LEMMA 3.4: (a) Laten $X_1 = (Y_1,Z_1)\ldots,X_N = (Y_N,Z_N)$ o.o. stochastische vectoren, ieder met continue tweedimensionale verdelingsfunctie F, $F = H.K$, $H,K \in \mathcal{F}_C$, en kansverdeling P op $(\mathbb{R}^2,\mathcal{B}^2)$, zijn. Dan geldt

$$P^N(R=r) = P^N(S=r) = \frac{1}{N!}$$

voor alle $r = (r_1,\ldots,r_N) \in \mathcal{R}$.

(b) $Y^{(\cdot)}$, $Z^{(\cdot)}$, R en S zijn onderling onafhankelijk.

(c) Indien bovendien P absoluut continu is t.o.v. een σ-eindige maat $\mu = \mu_1 \times \mu_2$ op $(\mathbb{R}^2,\mathcal{B}^2)$, dan worden de simultane dichtheden van $Y^{(1)},\ldots,Y^{(N)}$ en van $Z^{(1)},\ldots,Z^{(N)}$ respectievelijk t.o.v. de productmaten μ_{1N} en μ_{2N} gegeven door

$$\begin{cases} N!\ \Pi_{n=1}^N\ h(y^{(n)}) & \text{voor } y^{(1)} < \ldots < y^{(N)} \\ 0 & \text{anders} \end{cases}$$

en

$$\begin{cases} N!\ \Pi_{n=1}^N\ k(z^{(n)}) & \text{voor } z^{(1)} < \ldots < z^{(N)} \\ 0 & \text{anders} \end{cases}$$

waarbij h(k) de dichtheid van H(K) t.o.v. $\mu_1(\mu_2)$ voorstelt.

BEWIJS:
Het bewijs volgt direct uit lemma 3.1 en de onafhankelijkheid van de vectoren $(Y_1,\ldots,Y_N)$ en $(Z_1,\ldots,Z_N)$. □

We zullen vervolgens de theorie der rangtoetsen in het bijzonder verder ontwikkelen voor rangtoetsen voor de hypothese H_0 van isomorie. We beginnen met vast te stellen dat het kritieke gebied van een niet-gerandomiseerde rangtoets, voor het toetsen van H_0, een vereniging is van verzamelingen van de vorm

$$\{R = r\} = \{(x_1,\ldots,x_N) : x_n = x^{(r_n)},\ n = 1,\ldots,N\}.$$

We merken verder op dat we steeds zullen veronderstellen dat de verdelingsfuncties $F_1,\ldots,F_N$ van $X_1,\ldots,X_N$ allen continu zijn, zodat we ons zonder bezwaar kunnen beperken tot punten $x \in \mathbb{R}^N$ met coördinaten $(x_1,\ldots,x_N)$ waarvoor $x^{(1)} < \ldots < x^{(N)}$. Verder zullen we met H en K verzamelingen van N-tupels $(F_1,\ldots,F_N)$, met $F_n \in \mathcal{F}_C$, n = 1,...,N voorstellen. We schrijven dik-

wijls $P^N \in H(K)$ wanneer we bedoelen dat het bijbehorende N-tupel $(F_1,\ldots,F_N)$ tot H(K) behoort.

Volgens lemma 3.1.(a) geldt voor een kansverdeling $P^N \in H_0$ dat

$$P^N(R=r) = \frac{1}{N!}$$

voor alle $r \in \mathcal{R}$. Dit betekent dat voor een rangtoets voor H_0, de kans op verwerpen van H_0, indien H_0 juist is, onafhankelijk is van de keuze van $P^N \in H_0$. Een toets met deze eigenschap heet *verdelingsvrij*. Voor het toetsen van H_0 is iedere rangtoets voor H_0 verdelingsvrij.

OPMERKING 3.2: Men kan laten zien (zie LEHMANN [20], p.184 en HÁJEK & ŠIDÁK [12], p.43) dat voor het toetsen van H_0 de klasse der verdelingsvrije toetsen samenvalt met de klasse der permutatietoetsen. Rangtoetsen vormen een deelklasse van de klasse der permutatietoetsen welke zowel om praktische als ook om theoretische redenen van bijzondere betekenis is. Voor het toetsen van H_1 en H_2 gelden analoge beweringen.

In de volgende stelling construeren we de meest onderscheidende rangtoets voor het toetsen van H_0 tegen een enkelvoudige alternatieve hypothese. Onder het *(eindige) toetsingsprobleem* $(\alpha;H,K)$ verstaan we het probleem $P^N \in H$ tegen $P^N \in K$ te toetsen met onbetrouwbaarheidsdrempel α. Hierbij stellen H en K niet-lege disjuncte deelverzamelingen van de collectie van alle N-tupels $(F_1,\ldots,F_N)$, met $F_n \in \mathcal{F}_C$, $n = 1,\ldots,N$ voor. We noemen α de *onbetrouwbaarheidsdrempel* van het toetsingsprobleem $(\alpha;H,K)$ en ψ een *toets voor dit probleem*, indien

$$\int \psi dP^N \leq \alpha$$

voor iedere $P^N \in H$. In hoofdstuk 2 kwam reeds het (eindige) toetsingsprobleem $(\alpha_N;\theta(N),\tau(N))$ ter sprake, waarbij een enkelvoudige nulhypothese $\theta(N)$ tegen een enkelvoudige alternatieve hypothese $\tau(N)$ wordt getoetst.

STELLING 3.5: De meest onderscheidende rangtoets voor het (eindige) toet-

singsprobleem $(\alpha;H_0,K_0)$, waarbij voor gegeven willekeurige continue verdelingsfuncties $G_1,\ldots,G_N$,

$$K_0 = \{(F_1,\ldots,F_N) : F_n = G_n,\ n = 1,\ldots,N\}$$

een willekeurige enkelvoudige alternatieve hypothese voorstelt, wordt gegeven door

$$\begin{cases} = 1 & > k \\ \phi(r) = \gamma & \text{als } Q^N(R=r) = k \\ = 0 & < k, \end{cases}$$

waarbij Q^N de bij $G_1,\ldots,G_N$ behorende productmaat op $(\mathbb{R}^N,\mathcal{B}^N)$ voorstelt; de constanten k en γ zijn vastgelegd door $E\phi(R) = \alpha$.

BEWIJS:
Zij $\mathcal{B}_0^N$ een deel-σ-algebra van $\mathcal{B}^N$, namelijk de kleinste σ-algebra die alle verzamelingen $\{R=r\}$, $r \in \mathcal{R}$, bevat, en zij ν^N de telmaat op $(\mathbb{R}^N,\mathcal{B}_0^N)$ gedefinieerd door $\nu^N(R=r) = 1$ voor alle $r \in \mathcal{R}$. Dan is de restrictie van Q^N tot $\mathcal{B}_0^N$ absoluut continu t.o.v. de telmaat ν^N, met als dichtheid $q^N(x_1,\ldots,x_N) = Q^N(R=r)$ voor alle $(x_1,\ldots,x_N)$ met $x_n = x^{(r_n)}$, $n = 1,\ldots,N$ omdat geldt

$$Q^N(A) = \int_A q^N d\nu^N \qquad \text{voor alle } A \in \mathcal{B}_0^N.$$

Natuurlijk levert definitie van $p^N(x_1,\ldots,x_N) = 1/N!$ voor alle $(x_1,\ldots,x_N)$ met $x_n = x^{(r_n)}$, $n = 1,\ldots,N$, op volstrekt analoge wijze dat $P^N(A) = \int_A p^N d\nu^N$ voor alle $A \in \mathcal{B}_0^N$.

Hieruit en uit het feit dat iedere rangtoets voor H_0 verdelingsvrij is volgt onmiddellijk dat de meest onderscheidende rangtoets voor het toetsen van H_0 tegen een willekeurige enkelvoudige alternatieve hypothese K_0, volgens het lemma van Neyman-Pearson, H_0 verwerpt voor grote waarden van q^N/p^N, d.w.z. voor grote waarden van $Q^N(R=r)$. Hiermee is het bewijs voltooid. □

Daar, zoals uit stelling 3.5 blijkt, de meest onderscheidende rang-

toets voor H_0 afhankelijk is van de specifieke enkelvoudige alternatieve hypothese K_0 waartegen H_0 moet worden getoetst, zal een uniform meest onderscheidende rangtoets voor een samengestelde alternatieve hypothese in het algemeen niet bestaan. We zullen echter laten zien dat het mogelijk is rangtoetsen te construeren die "lokaal meest onderscheidend" zijn voor het toetsen van H_0.

Zij $\{G_\zeta, \zeta \in T\}$ een familie van continue verdelingsfuncties op $(\mathbb{R},\mathcal{B})$, en wel zodanig dat $G_\theta \neq G_\tau$ als $\theta \neq \tau$ $(\theta,\tau\in T)$. We zullen ons in dit hoofdstuk beperken tot het bijzondere geval dat T een *interval* in $\mathbb{R}$ is. Verder stelt θ_0 steeds een inwendig punt van T voor.

We introduceren thans allereerst de enkelvoudige alternatieve hypothese $K_{c,\Delta}$. Voor gegeven continue verdelingsfuncties $G_{\tau_1},\ldots,G_{\tau_N}$, met $\tau_1,\ldots,\tau_N \in T$, gegeven reële getallen $c_1,\ldots,c_N$ (welke niet allen gelijk zijn), een vast inwendig punt θ_0 van T, en een gegeven $\Delta \neq 0$, definiëren we

$$K_{c,\Delta} = \{(F_1,\ldots,F_N) : F_n = G_{\tau_n},\ \tau_n = \theta_0+\Delta c_n,\ n = 1,\ldots,N\}.$$

Vervolgens definiëren we de samengestelde alternatieve hypothese K_c (het *regressie-alternatief*) door

$$K_c = \bigcup_{\Delta>0} K_{c,\Delta}.$$

De getallen $c_1,\ldots,c_N$ noemen we de *regressieconstanten*.

We beschouwen nu het toetsingsprobleem $(\alpha;H_0,K_c)$, d.w.z. het toetsen van de hypothese H_0 van isomorie tegen het regressie-alternatief K_c. Zoals we nog zullen zien kunnen verschillende interessante toetsingsproblemen als speciale gevallen van het toetsingsprobleem $(\alpha;H_0,K_c)$ verkregen worden.

DEFINITIE 3.1: Een rangtoets heet een *lokaal meest onderscheidende rangtoets* voor het toetsingsprobleem $(\alpha;H_0,K_c)$, indien er een $\varepsilon > 0$ bestaat, zódat de toets uniform meest onderscheidend is voor het toetsingsprobleem $(\alpha;H_0,K_c(\varepsilon))$, waarbij

$$K_c(\varepsilon) = \bigcup_{0<\Delta<\varepsilon} K_{c,\Delta},$$

in de klasse van alle rangtoetsen met dezelfde onbetrouwbaarheidsdrempel.

Alvorens nu over te gaan tot de constructie van een lokaal meest onderscheidende rangtoets voor het toetsingsprobleem $(\alpha;H_0,K_c)$ vermelden wij nog een technisch lemma.

LEMMA 3.6: Zij (Ω,A) een meetbare ruimte, voorzien van een σ-eindige maat μ. Zij $\{h_N\}$ een rij A-meetbare functies, welke μ-b.o. convergeert naar een A-meetbare functie h. Als

$$\text{limsup}_{N\to\infty} \int |h_N| d\mu \leq \int |h| d\mu < \infty,$$

dan geldt

$$\lim_{N\to\infty} \int_A h_N d\mu = \int_A h d\mu$$

voor iedere $A \in A$.

BEWIJS:
Zie HÁJEK & ŠIDÁK, p.64. □

We introduceren verder nog enige begrippen en notaties. Laat, voor iedere $\zeta \in T$, de kansmaat behorende bij $G_\zeta(x)$ absoluut continu zijn t.o.v. een σ-eindige maat μ op $(\mathbb{R},\mathcal{B})$, met dichtheid $g_\zeta(x)$. Verder zullen we veronderstellen dat de afgeleide van $g_\zeta(x)$ m.b.t. ζ in het punt $\zeta = \theta_0$ voor μ-bijna alle x bestaat. Evenals in voorwaarde (I), hoofdstuk 2, voeren we voor de partiële afgeleide van $g_\zeta(x)$ m.b.t. ζ in het punt $\zeta = \theta_0$ de speciale notatie $\dot{g}_0(x)$ in. Verder noteren we $G_0(x)$ voor $G_{\theta_0}(x)$, $g_0(x)$ voor $g_{\theta_0}(x)$, en $G_0^{-1}(t)$ voor $G_{\theta_0}^{-1}(t)$ $(0<t<1)$. Hierbij stelt $G_{\theta_0}^{-1}(t)$ de inverse van G_{θ_0} voor: $G_{\theta_0}^{-1}(t) = \inf\{x : G_{\theta_0}(x) \geq t\}$ voor alle $0 < t < 1$.

DEFINITIE 3.2: Zij γ een sommeerbare functie op $(0,1)$. Dan noemen we de ge-

tallen

$$(3.5) \qquad a_N(n,\gamma) = E\gamma(U^{(n)}), \qquad n = 1,\ldots,N,$$

de *exacte scores* behorende bij de functie γ. Hierbij stelt $(U^{(1)},\ldots,U^{(N)})$ een geordende steekproef van omvang N uit een homogene verdeling op (0,1) voor. De functie γ noemt men wel de *scorefunctie*.

In het vervolg zullen scorefuncties van de gedaante

$$(3.6) \qquad \gamma_0(t) = \frac{\dot{g}_0(G_0^{-1}(t))}{g_0(G_0^{-1}(t))}, \qquad 0 < t < 1,$$

van bijzonder belang blijken te zijn. Dit zijn scorefuncties behorende bij een familie verdelingsfuncties $\{G_\zeta,\ \zeta \in T\}$ en een zeker inwendig punt θ_0 van T. We merken hierbij nog op dat voor de bij (3.6) behorende exacte scores geldt dat

$$(3.7) \qquad a_N(n,\gamma_0) = E_V \frac{\dot{g}_0(V^{(n)})}{g_0(V^{(n)})}, \qquad n = 1,\ldots,N,$$

waarbij $(V^{(1)},\ldots,V^{(N)})$ een geordende steekproef van omvang N uit de kansverdeling met verdelingsfunctie G_0 voorstelt.

Onder zekere omstandigheden kunnen de exacte scores benaderd worden door getallen die dikwijls eenvoudiger te berekenen zijn. We kunnen bijvoorbeeld $a_N(n,\gamma)$ (zie definitie 3.2) als volgt benaderen:

$$(3.8) \qquad a_N(n,\gamma) \simeq \gamma(E(U^{(n)})) =$$

$$= \gamma(\tfrac{n}{N+1}), \qquad n = 1,\ldots,N.$$

We noemen de getallen $\gamma(\frac{n}{N+1})$, $n = 1,\ldots,N$ *benaderde scores* behorende bij de scorefunctie γ.

In de volgende stelling zullen we de lokaal meest onderscheidende rangtoets voor het toetsingsprobleem $(\alpha;H_0,K_c)$ construeren. Aangezien de

voorwaarde waaronder wij deze stelling zullen bewijzen ook in het vervolg een rol speelt, zullen wij deze apart vermelden. Met het oog op de constructie van lokaal meest onderscheidende rangtoetsen voor andere toetsingsproblemen wordt deze voorwaarde in een wat algemenere vorm gegeven dan voor de volgende stelling noodzakelijk is.

VOORWAARDE (II): Zij θ_0 een inwendig punt van T. De kansmaten behorende bij de familie $\{G_\zeta, \zeta \in T\}$ van k-dimensionale verdelingsfuncties worden gedomineerd door een σ-eindige maat μ op $(\mathbb{R}^k, \mathcal{B}^k)$. De dichtheid van de kansmaat behorende bij G_ζ ten opzichte van μ noteren we met g_ζ. Er is een verzameling $N_0 \in \mathcal{B}^k$ met $\mu(N_0) = 0$, zodat (in de eerder vermelde notatie) $g_0(x) > 0$ en $\dot{g}_0(x)$ bestaat voor alle $x \in N_0^C$.
Tenslotte geldt

$$(3.9) \qquad \lim_{\zeta\to\theta_0} \int \left|\frac{g_\zeta - g_0}{\zeta-\theta_0}\right| d\mu = \int \lim_{\zeta\to\theta_0} \left|\frac{g_\zeta - g_0}{\zeta-\theta_0}\right| d\mu = \int |\dot{g}_0| d\mu < \infty.$$

STELLING 3.7: Zij k = 1 en θ_0 een inwendig punt van T. Laat de familie $\{G_\zeta, \zeta \in T\}$ aan voorwaarde (II) voldoen en zij $\gamma_0 = \dot{g}_0(G_0^{-1})/g_0(G_0^{-1})$. Dan wordt de lokaal meest onderscheidende rangtoets voor het toetsingsprobleem $(\alpha; H_0, K_c)$ gegeven door

$$\begin{cases} \phi(r) = 1 & \text{als } T \geq c_\alpha, \\ = 0 & \text{als } T < c_\alpha, \end{cases}$$

waarbij $T = \sum_{n=1}^N c_n a_N(r_n, \gamma_0)$ en $r_1, \ldots, r_n$ de waarden zijn der rangnummers van $X_1, \ldots, X_N$, en waarin de onbetrouwbaarheidsdrempel α een der mogelijke waarden

$$P^N\Big(\sum_{n=1}^N c_n a_N(R_n, \gamma_0) \geq c\Big) \qquad c \in \mathbb{R},$$

$P^N \in H_0$, voorstelt.

BEWIJS:
De productmaat behorende bij de N-dimensionale verdelingsfunctie $\Pi_{n=1}^N G_{\tau_n}$,

met $\tau_n = \theta_0+\Delta c_n$, $n = 1,\ldots,N$ zullen we noteren met Q_Δ^N. Dan geldt

$$Q_\Delta^N(R=r) = \int_{R=r} \Pi_{n=1}^N g_{\theta_0+\Delta c_n}(x_n)d\mu_N(x_1,\ldots,x_N) =$$

$$= \int_{R=r} \Pi_{n=1}^N g_0(x_n)d\mu_N(x_1,\ldots,x_N) +$$

$$+ \int_{R=r} (\Pi_{n=1}^N g_{\theta_0+\Delta c_n}(x_n) - \Pi_{n=1}^N g_0(x_n))d\mu_N(x_1,\ldots,x_N) =$$

$$= \frac{1}{N!} + \Delta \sum_{k=1}^N \int_{R=r} \frac{g_{\theta_0+\Delta c_k}(x_k) - g_0(x_k)}{\Delta} \cdot \Pi_{i=k+1}^N g_0(x_i) \cdot$$

$$\cdot \Pi_{j=1}^{k-1} g_{\theta_0+\Delta c_j}(x_j)d\mu_N(x_1,\ldots,x_n),$$

waarbij de derde overgang gebruik maakt van lemma 3.1.

Uit voorwaarde (II) volgt nu dat

$$\lim_{\Delta\to 0} \frac{g_{\theta_0+\Delta c_k}(x_k) - g_0(x_k)}{\Delta} \Pi_{i=k+1}^N g_0(x_i) \Pi_{j=1}^{k-1} g_{\theta_0+\Delta c_j}(x_j) =$$

$$= c_k \dot{g}_0(x_k) \Pi_{\substack{i=1 \\ i\neq k}}^N g_0(x_i), \qquad k = 1,\ldots,N,$$

geldt voor μ_N- bijna alle $(x_1,\ldots,x_N)$. Verder impliceert (3.9) dat

$$\lim_{\Delta\to 0} \int \left| \frac{g_{\theta_0+\Delta c_k}(x_k) - g_0(x_k)}{\Delta} \right| \Pi_{i=k+1}^N g_0(x_i) \Pi_{j=1}^{k-1} g_{\theta_0+\Delta c_j}(x_j)d\mu_N(x_1,\ldots,x_N) =$$

$$= \int \lim_{\Delta\to 0} \left| \frac{g_{\theta_0+\Delta c_k}(x_k) - g_0(x_k)}{\Delta} \right| d\mu(x_k) =$$

$$= |c_k| \cdot \int |\dot{g}_0| d\mu < \infty$$

Maar hieruit volgt wegens lemma 3.6 dat

$$\lim_{\Delta\to 0} \sum_{k=1}^{N} \int_{R=r} \frac{g_{\theta_0+\Delta c_k}(x_k) - g_0(x_k)}{\Delta} \prod_{i=k+1}^{N} g_0(x_i) \prod_{j=1}^{k-1} g_{\theta_0+\Delta c_j}(x_j) d\mu_N(x_1,\ldots,x_N) =$$

$$= \sum_{k=1}^{N} \int_{R=r} c_k \dot{g}_0(x_k) \prod_{\substack{i=1 \\ i\neq k}}^{N} g_0(x_i) d\mu_N(x_1,\ldots,x_N).$$

We merken nu op dat

$$\sum_{k=1}^{N} \int_{R=r} c_k \dot{g}_0(x_k) \prod_{\substack{i=1 \\ i\neq k}}^{N} g_0(x_i) d\mu_N(x_1,\ldots,x_N) =$$

$$= \sum_{k=1}^{N} \int_{R=r} c_k \frac{\dot{g}_0(x_k)}{g_0(x_k)} \prod_{i=1}^{N} g_0(x_i) d\mu_N(x_1,\ldots,x_N) =$$

$$= \frac{1}{N!} \sum_{k=1}^{N} c_k E_V^{2)} \left(\frac{\dot{g}_0(V_k)}{g_0(V_k)} \mid R=r\right) =$$

$$= \frac{1}{N!} \sum_{k=1}^{N} c_k a_N(r_k,\gamma_0),$$

waarbij $V_1,\ldots,V_N$ een steekproef uit de kansverdeling met verdelingsfunctie G_0 voorstelt. De eerste overgang maakt gebruik van voorwaarde (II), terwijl de tweede en derde overgang volgen door toepassing van lemma 3.1 en 3.2.

Hieruit kunnen we nu concluderen dat voor voldoend kleine $\varepsilon > 0$

$$\sum_{n=1}^{N} c_n a_N(r_n,\gamma_0) > \sum_{n=1}^{N} c_n a_N(r_n',\gamma_0)$$

impliceert dat

$$Q_\Delta^N(R=r) > Q_\Delta^N(R=r')$$

---

2) E_V stelt de verwachting m.b.t. de kansverdeling van de stochastische vector $V = (V_1,\ldots,V_N)$ voor.

voor alle $0 < \Delta < \varepsilon$. Maar dit betekent dat de toets die H_0 verwerpt voor grote waarden van $T = \sum_{n=1}^{N} c_n a_N(R_n,\gamma_0)$, op grond van stelling 3.5, het onderscheidingsvermogen maximaliseert voor het samengestelde alternatief $K_c(\varepsilon)$, in de klasse van alle rangtoetsen met dezelfde onbetrouwbaarheidsdrempel. □

OPMERKING 3.3: De lokaal meest onderscheidende rangtoetsen welke in stelling 3.7 gegeven worden zijn niet-gerandomiseerde toetsen. Toetsen, gebaseerd op rangtoetsingsgrootheden $T = \sum_{n=1}^{N} c_n a_N(R_n,\gamma_0)$, zijn lokaal meest onderscheidende rangtoetsen met onbetrouwbaarheden α_0 van de vorm $\alpha_0 = P^N(T \geq c)$. Deze waarden van α_0 zijn de zgn. *natuurlijke onbetrouwbaarheidsdrempels* van de toets gebaseerd op $T = \sum_{n=1}^{N} c_n a_N(R_n,\gamma_0)$. Onder een sterkere voorwaarde dan voorwaarde II (zie WITTING & NÖLLE [24] p.127) kunnen ook lokaal meest onderscheidende rangtoetsen voor andere waarden van α_0 bepaald worden. De bijbehorende rangtoetsingsgrootheden zijn dan echter niet meer van de eenvoudige gedaante $T = \sum_{n=1}^{N} c_n a_N(R_n,\gamma_0)$ en blijven hier buiten beschouwing.

OPMERKING 3.4: In het algemeen zullen de scores $a_N(n,\gamma_0)$, $n = 1,\ldots,N$ niet alleen van de familie verdelingsfuncties $\{G_\zeta, \zeta \in T\}$ maar ook nog van het inwendige punt $\theta_0 \in T$ afhangen. Deze afhankelijkheid van θ_0 kan echter in een tweetal speciale gevallen geëlimineerd worden:

(i) ζ is een *lokatieparameter*, d.w.z.

$$G_\zeta(x) = G(x-\zeta), \qquad -\infty < x < \infty,\ \zeta \in T, \tag{3.10}$$

waarin G een vaste verdelingsfunctie is. In dit geval vinden we

$$\gamma_0(t) = \frac{\dot{g}_0(G_0^{-1}(t))}{g_0(G_0^{-1}(t))} = -\frac{g'(G^{-1}(t))}{g(G^{-1}(t))}, \qquad 0 < t < 1, \tag{3.11}$$

waarbij g' de afgeleide (naar x) van g voorstelt. Dit verifeert men eenvoudig door op te merken dat, op grond van (3.10), $\dot{g}_0(x) = -g'(x-\theta_0)$ en $G_0^{-1}(t) = G^{-1}(t) + \theta_0$, $0 < t < 1$. We merken verder nog op dat in het geval dat

(3.12) $\quad G_{\zeta,\sigma}(x) = G(\frac{x-\zeta}{\sigma}), \qquad -\infty < x < \infty,\ \zeta \in T,$

voor zekere verdelingsfunctie G en zekere $\sigma > 0$, de scores behorende bij de familie verdelingsfuncties $\{G_{\zeta,\sigma}, \zeta \in T\}$ en een inwendig punt $\theta_0 \in T$, op vermenigvuldiging met de positieve constante σ^{-1} na, gelijk zijn aan de scores $a_N(n,\gamma_0)$ behorende bij de scorefunctie (3.11). Maar dit betekent dat de lokaal meest onderscheidende rangtoets voor het toetsen van H_0 tegen een regressie-alternatief K_c, waarbij de daarin optredende verdelingsfuncties $G_{\tau_1},\ldots,G_{\tau_N}$ de gedaante (3.10) hebben, in feite alleen van het type van de verdelingsfunctie G afhangt.

(ii) Beschouw het geval dat

(3.13) $\quad G_\zeta(x) = G(xe^{-\zeta}) \qquad -\infty < x < \infty,\ \zeta \in T.$

Deze familie van verdelingsfuncties is equivalent met het meer gebruikelijke model $G_\sigma(x) = G(x/\sigma)$, $-\infty < x < \infty$, met schaalparameter $\sigma = e^\zeta > 0$. De schrijfwijze met parameter ζ is echter voor ons doel meer geschikt. Men berekent eenvoudig dat

(3.14) $$\gamma_0(t) = -1 - G_0^{-1}(t)\,\frac{\dot g_0(G_0^{-1}(t))}{g_0(G_0^{-1}(t))}, \qquad 0 < t < 1.$$

Men verifieert nu zonder veel moeite dat de scorefunctie (3.14) niet van de parameter θ_0 afhangt.

OPMERKING 3.5: Omdat we ons bij de definitie van K_c beperken tot waarden $\Delta > 0$ spreken we van een "eenzijdig" toetsingsprobleem $(\alpha;H_0,K_c)$. Het geval dat $\Delta < 0$ (in plaats van $\Delta > 0$) is volstrekt analoog en blijft hier buiten beschouwing. In de practijk wenst men dikwijls H_0 te toetsen tegen een "tweezijdig" alternatief $\bigcup_{\Delta\neq 0} K_{c,\Delta}$. Men verwerpt dan H_0 ten gunste van het tweezijdige alternatief voor grote waarden van $|T|$. De bijbehorende rangtoets, gebaseerd op $|T|$, is dan echter geen lokaal meest onderscheidende rangtoets meer.

In stelling 3.7 werden lokaal meest onderscheidende rangtoetsen bepaald voor het toetsen van de hypothese H_0 van isomorie tegen een regressie-alternatief K_c. Door speciale keuze van de regressieconstanten $c_1,\ldots,c_N$ en de verdelingsfuncties $G_{\tau_1},\ldots,G_{\tau_N}$ kunnen belangrijke bijzondere gevallen verkregen worden.

In het bijzonder belangrijke speciale geval dat

$$(3.15)\qquad \begin{cases} c_n = 1 & \text{als } n = 1,\ldots,N_1 \\ c_n = 0 & \text{als } n = N_1+1,\ldots,N_1+N_2, \end{cases}$$

waarbij $N_1 + N_2 = N$, spreken we van een *twee-steekproeven probleem*. We hebben dan immers te maken met stochastische grootheden $X_1,\ldots,X_N$, waarbij $X_1,\ldots,X_{N_1}$, (de "eerste" steekproef) afkomstig zijn uit éénzelfde kansverdeling, terwijl de stochastische grootheden $X_{N_1+1},\ldots,X_{N_1+N_2}$ (de "tweede" steekproef) afkomstig zijn uit een tweede kansverdeling.

De nulhypothese H_0 houdt nu in dat beide steekproeven, die onderling onafhankelijk zijn, afkomstig zijn uit éénzelfde kansverdeling. Door keuze van de in het alternatief optredende verdelingsfuncties G_{θ_0} en $G_{\theta_0+\Delta}$ $(\Delta>0)$ kunnen twee-steekproeven toetsen verkregen worden, welke lokaal meest onderscheidende rangtoetsen zijn voor het toetsen van H_0 tegen het desbetreffende alternatief. We schrijven verder G_0 en G_Δ in plaats van G_{θ_0} en $G_{\theta_0+\Delta}$.

We behandelen nu een paar interessante gevallen.

H_0 tegen verschuivingsalternatieven. Zij G een verdelingsfunctie met een dichtheid g t.o.v. de lebesguemaat λ en zij

$$(3.16)\qquad G_\Delta(x) = G(x-\Delta) \qquad -\infty < x < \infty,\ \Delta > 0,$$

een familie van verschuivingsalternatieven. Dan volgt uit (3.15), stelling 3.7 en opmerking 3.4.(i) dat, onder voorwaarde (II) met $k = 1$ en $\mu = \lambda$, de toets die H_0 verwerpt voor grote waarden van

$$(3.17) \qquad T = \sum_{n=1}^{N_1} a_N(R_n,\gamma_0) = -\sum_{n=1}^{N_1} E_V \frac{g'(V^{(R_n)})}{g(V^{(R_n)})},$$

waarbij $(V^{(1)},\ldots,V^{(N)})$ een geordende steekproef van omvang N uit de kansverdeling met verdelingsfunctie G voorstelt, de lokaal meest onderscheidende twee-steekproeven rangtoets is voor het toetsen van H_0 tegen de alternatieve hypothese dat de "eerste" steekproef $X_1,\ldots,X_{N_1}$ naar rechts verschoven is t.o.v. de "tweede" steekproef $X_{N_1+1},\ldots,X_{N_1+N_2}$: X_n heeft ververdelingsfunctie $G(x-\Delta)$ voor $n = 1,\ldots,N_1$ en verdelingsfunctie $G(x)$ voor $n = N_1+1,\ldots,N_1+N_2$.

Het is nu eenvoudig in te zien dat een aantal bekende rangtoetsen voor het twee-steekproeven probleem, tegen verschuivingsalternatieven, als lokaal meest onderscheidende rangtoetsen verkregen kunnen worden.

VOORBEELD 3.1: Voor het bepalen van de lokaal meest onderscheidende rangtoets voor het toetsen van H_0 tegen normale verschuivingsalternatieven, mogen we, zoals men eenvoudig nagaat (zie opmerking 3.4), zonder bezwaar voor G de standaard normale verdelingsfunctie Φ kiezen. Dit betekent dat

$$g(x) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}x^2}, \quad g'(x) = -xg(x), \qquad -\infty < x < \infty,$$

zodat, op grond van (3.17), de lokaal meest onderscheidende rangtoets H_0 verwerpt voor grote waarden van de toetsingsgrootheid $T = \sum_{n=1}^{N_1} E_V(V^{(R_n)})$, waarbij $R_1,\ldots,R_{N_1}$ de rangnummers van $X_1,\ldots,X_{N_1}$ onder $X_1,\ldots,X_N$ voorstellen, en $(V^{(1)},\ldots,V^{(N)})$ een geordende steekproef van omvang N uit de standaard normale verdeling is.

Dit is de *normale scores* toets van *Fisher-Yates*. De verwachtingen van geordende waarnemingen uit een standaard normale verdeling zijn voor $N \leq 50$ getabelleerd; voor zekere waarden van α zijn kritieke waarden berekend.

OPMERKING 3.6: Een modificatie van de toets van Fisher-Yates is afkomstig

van *Van der Waerden*. Deze koos als toetsingsgrootheid $\sum_{n=1}^{N_1} \Phi^{-1}(\frac{R_n}{N+1})$, waarbij Φ^{-1} de inverse van de standaard normale verdelingsfunctie voorstelt. De exacte scores $E_V(V^{(n)})$ van de toets van Fisher-Yates worden vervangen door de benaderde scores $\Phi^{-1}(\frac{n}{N+1})$ (zie (3.8)). Zoals nog aangetoond zal worden zijn de toetsen van Fisher-Yates en Van der Waerden asymptotisch aequivalent.

VOORBEELD 3.2: De twee-steekproeven toets van *Wilcoxon* is de lokaal meest onderscheidende rangtoets voor het toetsen van H_0 tegen logistieke verschuivingsalternatieven. Om dit in te zien merken we op dat we zonder bezwaar (zie opmerking 3.4) voor G de standaard logistieke verdelingsfunctie F kunnen kiezen. De verdelingsfunctie van de standaard logistieke verdeling wordt gegeven door

$$F(x) = \frac{1}{1 + e^{-x}}, \qquad -\infty < x < \infty.$$

Men verifieert eenvoudig dat indien $G = F$, $-\frac{g'}{g} = 2F - 1$, zodat de lokaal meest onderscheidende rangtoets H_0 verwerpt voor grote waarden van de toetsingsgrootheid $\sum_{n=1}^{N_1} E_V F(V^{(R_n)})$, waarbij $R_1,\ldots,R_{N_1}$ de rangnummers van $X_1,\ldots,X_{N_1}$ onder $X_1,\ldots,X_N$ voorstellen en $(V^{(1)},\ldots,V^{(N)})$ een geordende steekproef van omvang N uit de standaard logistieke verdeling is. Omdat F continu is, is $F(V^{(n)})$ in feite de n^{de}-geordende waarneming van een steekproef van de omvang N uit de homogene verdeling op (0,1) zodat $E_V F(V^{(n)}) = \frac{n}{N+1}$, m.a.w. we verwerpen H_0 voor grote waarden van de toetsingsgrootheid $\sum_{n=1}^{N_1} R_n$. Deze toets is de bekende twee-steekproeven toets van Wilcoxon.

OPMERKING 3.7: Men gaat eenvoudig na dat normale en logistieke verschuivingsalternatieven aan voorwaarde (II), met $k = 1$, $N_0^c = \mathbb{R}$ en $\mu = \lambda$, van stelling 3.7 voldoen.

Meer in het algemeen kan men m.b.v. stelling 3.7 het volgende resultaat voor het twee-steekproeven probleem, tegen verschuivingsalternatieven, bewijzen.

GEVOLG 3.8: Zij G een verdelingsfunctie met een positieve continue differentieerbare λ-dichtheid g (met afgeleide g'), waarvoor $\int |g'| d\lambda < \infty$. Zij verder $\gamma_0 = -\frac{g'(G^{-1})}{g(G^{-1})}$. Dan wordt de lokaal meest onderscheidende rangtoets voor het twee-steekproeven probleem, waarbij H_0 getoetst wordt tegen verschuivings alternatieven (3.16), gegeven door

$$\begin{cases} \phi(r) = 1 & \text{als } \sum_{n=1}^{N_1} a_N(r_n,\gamma_0) \geq c \\ = 0 & \text{als } \sum_{n=1}^{N_1} a_N(r_n,\gamma_0) < c \end{cases}$$

De onbetrouwbaarheid α_0 van de toets wordt gegeven door

$$\alpha_0 = P^N(\sum_{n=1}^{N_1} a_N(R_n,\gamma_0) \geq c)$$

voor iedere $P^N \in H_0$.

BEWIJS:

Het bewijs volgt uit stelling 3.7 en het feit dat, wegens (3.16), de gegeven voorwaarden voor de verdelingsfunctie G, voorwaarde (II), met k = 1, $N_0^c = \mathbb{R}$ en $\mu = \lambda$, impliceren. Om dit in te zien beginnen we met op te merken dat, op grond van (3.16), $g_\zeta(x) = g(x-\zeta)$, $-\infty < x < \infty$, $\zeta \in T$. Hieruit volgt onmiddellijk dat $\dot{g}_0 = -g'$.

We gaan aantonen dat de voorwaarden voor de verdelingsfunctie G relatie (3.9) van voorwaarde (II) impliceren. Daartoe merken we allereerst op dat

$$\int \left| \frac{g(x-\Delta)-g(x)}{\Delta} \right| dx = \int \left| \frac{1}{\Delta} \int \chi_{[x-\Delta,x]}(y) g'(y) dy \right| dx.$$

Het bewijs berust nu verder op toepassing van lemma 3.6. Zij Δ_N, N = 1,2,... een rij positieve getallen waarvoor $\lim_{N\to\infty} \Delta_N = 0$. Omdat G een continue differentieerbare dichtheid g (met afgeleide g') bezit geldt dat

$$\lim_{\Delta_N \to 0} \frac{1}{\Delta_N} \int \chi_{[x-\Delta_N,x]}\, g'(y)\, dy = g'(x), \qquad -\infty < x < \infty.$$

Verder is

$$\limsup_{\Delta_N \to 0} \int \left| \frac{1}{\Delta_N} \int \chi_{[x-\Delta_N,x]}(y)\, g'(y) dy \right| dx \leq$$

$$\leq \limsup_{\Delta_N \to 0} \frac{1}{\Delta_N} \int \int \chi_{[x-\Delta_N,x]}(y)\, |g'(y)| dy dx =$$

$$= \limsup_{\Delta_N \to 0} \frac{1}{\Delta_N} \int |g'(y)| \int \chi_{[y,y+\Delta_N]}\, dx dy =$$

$$= \int |g'(y)| dy < \infty.$$

Toepassing van lemma 3.6 levert nu

$$\lim_{\Delta_N \to 0} \int \left| \frac{g(x-\Delta_N)-g(x)}{\Delta_N} \right| dx = \int |g'(y)| dy$$

voor iedere rij $\{\Delta_N\}$ met $\lim_{N \to \infty} \Delta_N = 0$. Hiermee is het bewijs voltooid. □

H_0 tegen Lehmann-alternatieven. Zij G een continue verdelingsfunctie, en zij

$$(3.18) \qquad G_\Delta(x) = (1-\Delta)G(x) + \Delta G^2(x), \qquad -\infty < x < \infty,\ 0 < \Delta < 1,$$

een familie van zgn. Lehmann-alternatieven. De verdelingsfuncties G_Δ, $0 < \Delta < 1$, hebben t.o.v. de kansmaat P, d.i. de kansmaat behorende bij de verdelingsfunctie G, P-dichtheden

$$(3.19) \qquad g_\Delta(x) = 1 + \Delta(2G(x)-1), \qquad -\infty < x < \infty,\ 0 < \Delta < 1.$$

Nu impliceert stelling 3.7, onder toepassing van voorwaarde (II) met k = 1 en $\mu = P$, dat de toets die H_0 verwerpt voor grote waarden van $\sum_{n=1}^{N_1}(2\frac{R_n}{N+1} - 1)$ de lokaal meest onderscheidende rangtoets is voor het toetsen van H_0 tegen de alternatieven (3.18) voor iedere keuze van G. Deze toets is equivalent met de twee-steekproeven toets van Wilcoxon.

Op volkomen analoge wijze kunnen lokaal meest onderscheidende rangtoetsen voor het toetsen van H_0 tegen schaalalternatieven geconstrueerd worden. Ook toetsen voor H_0 tegen verloopalternatieven kunnen, door speciale keuze van de regressieconstanten $c_1,\ldots,c_N$ als bijzonder geval van stelling 3.7 verkregen worden.

We besluiten onze uiteenzetting over de eindige theorie der rangtoetsen met het vermelden van enige resultaten met betrekking tot rangtoetsen voor de hypothesen H_1 en H_2.

In de eerste plaats merken we op dat $H_1 \subset H_0$ zodat iedere rangtoets voor H_0 in feite ook een toets voor H_1 is. Bij de constructie van optimale toetsen voor de hypothese H_1 zullen we echter gebruik moeten maken van het feit dat, onder H_1, de gemeenschappelijke verdeling van $X_1,\ldots,X_N$ symmetrisch is.

We beperken ons hier tot de constructie van de lokaal meest onderscheidende rangtoets voor de hypothese H_1 van symmetrie tegen verschuivingsalternatieven.

STELLING 3.9.: Zij G een symmetrische verdelingsfunctie met een symmetrie punt 0 en met een positieve continue-differentieerbare dichtheid g (met afgeleide g'), waarvoor $\int |g'|d\lambda < \infty$. Zij verder

$$\gamma_0^+(t) = -\frac{g'(G^{-1}(\frac{1}{2}+\frac{1}{2}t))}{g(G^{-1}(\frac{1}{2}+\frac{1}{2}t))}, \qquad 0 < t < 1,$$

en laten $a_N(n,\gamma_0^+)$, n=1,...,N, de scores behorende bij de scorefunctie γ_0^+ zijn. Dan wordt de lokaal meest onderscheidende rangtoets voor het probleem H_1 te toetsen tegen verschuivingsalternatieven (3.16), gegeven door

$$\begin{cases} \phi(r^+) = 1 & \text{als } \sum_{n=1}^{N} a_N(r_n^+,\gamma_0^+)\operatorname{sgn} x_n \geq c \\ \quad = 0 & \text{als } \sum_{n=1}^{N} a_N(r_n^+,\gamma_0^+)\operatorname{sgn} x_n < c, \end{cases}$$

waarbij $r^+ = (r_1^+,\ldots,r_N^+)$; $r_1^+,\ldots,r_N^+$ en $x_1,\ldots,x_N$ zijn respectievelijk de waarden van de rangnummers van $|X_1|,\ldots,|X_N|$ en van $X_1,\ldots,X_N$. De onbetrouwbaarheid α_0 van de toets wordt gegeven door

$$\alpha_0 = P^N(\sum_{n=1}^{N} a_N(R_n^+,\gamma_0^+)\operatorname{sgn} X_n \geq c)$$

voor iedere $P^N \in H_1$.

BEWIJS:
We beginnen met op te merken dat de in de stelling gegeven toets inderdaad een rangtoets voor H_1 is: de bijbehorende kritieke functie is een functie die slechts van de vectoren $R^+ = (R_1^+,\ldots,R_N^+)$ en sgn X = (sgn X_1,...,sgn X_N) afhangt.

Het bewijs van de stelling verloopt analoog aan dat van stelling 3.7. In plaats van lemma 3.1 maken we gebruik van lemma 3.3, terwijl stelling 3.5 vervangen wordt door een analoog resultaat, dat hier onvermeld blijft, voor het toetsen van H_1 tegen een enkelvoudig alternatief. □

Lokaal meest onderscheidende rangtoetsen voor de hypothese H_1, tegen verschuivingsalternatieven, kunnen verkregen worden door speciale keuze van de verdelingsfunctie G. Men gaat bijvoorbeeld eenvoudig na dat de éénsteekproef toets van *Wilcoxon* (symmetrie-toets van Wilcoxon) de lokaal meest onderscheidende rangtoets is voor het toetsen van de hypothese H_1 tegen verschuivingsalternatieven van het logistieke type. De zgn. *tekentoets*, gebaseerd op het aantal positieve waarden onder $X_1,\ldots,X_N$, is de lokaal meest onderscheidende rangtoets voor de hypothese H_1 tegen verschuivingsalternatieven van het dubbel-exponentiële type. Ook kunnen rangtoetsen voor H_1 tegen normale verschuivingsalternatieven van het Fisher-Yates en Van der Waerden type aangegeven worden.

Tenslotte zullen we, voor wat betreft het toetsen van de hypothese H_2 van onafhankelijkheid, volstaan met te laten zien dat de rangcorrelatie toets van *Spearman* de lokaal meest onderscheidende rangtoets is voor het toetsen van de hypothese H_2 tegen Lehmann-alternatieven van een zeker type.

STELLING 3.10: Zij k = 2 en zij

$$(3.20) \qquad G_\Delta(x) = H(y)K(z)(1+\Delta(H(y)-1)(K(z)-1)), \qquad x = (y,z),$$

$-\infty < y,z < \infty$, $0 < \Delta < 1$, $H,K \in F_c$. Dan wordt de lokaal meest onderscheidende rangtoets voor het probleem H_2 te toetsen tegen Lehmann-alternatieven (3.20), voor iedere keuze van $H,K \in F_c$, gegeven door

$$\begin{cases} \phi(r,s) = 1 & \text{als } \sum_{n=1}^{N} r_n s_n \geq c \\ \phi(r,s) = 0 & \text{als } \sum_{n=1}^{N} r_n s_n < c, \end{cases}$$

waarbij $r_1,\ldots,r_N$ en $s_1,\ldots,s_N$ de waarden der rangnummers van $Y_1,\ldots,Y_N$ respectievelijk $Z_1,\ldots,Z_N$ voorstellen. Dit is de rangcorrelatietoets van Spearman. De onbetrouwbaarheid α_0 van de toets wordt gegeven door

$$\alpha_0 = P^N(\textstyle\sum_{n=1}^{N} R_n S_n \geq c)$$

voor iedere $P^N \in H_2$

BEWIJS:
De in de stelling gegeven toets is inderdaad een rangtoets voor H_2: de bijbehorende kritieke functie is een functie die slechts van de vectoren $R = (R_1,\ldots,R_N)$ en $S = (S_1,\ldots,S_N)$ der rangnummers van $Y = (Y_1,\ldots,Y_N)$ respectievelijk van $Z = (Z_1,\ldots,Z_N)$ afhangt.

Het bewijs van de stelling verloopt analoog aan dat van stelling 3.7. In plaats van lemma 3.1 maken we gebruik van lemma 3.4, terwijl stelling 3.5 vervangen wordt door een analoog resultaat voor het toetsen van H_2 tegen een

enkelvoudig alternatief. We merken nog op dat aan voorwaarde (II), met k = 2 $N_0^c = \mathbb{R}^2$ en $\mu = P^2$, waarbij P^2 de kansmaat behorende bij de verdelingsfunctie G = H.K voorstelt, voldaan is. □

In het voorafgaande hebben we gezien dat voor zekere toetsingsproblemen uit de verdelingsvrije statistiek, lokaal meest onderscheidende rangtoetsen kunnen worden bepaald. Hoewel deze toetsen het onderscheidingsvermogen "lokaal" maximaliseren in de klasse van alle rangtoetsen met dezelfde onbetrouwbaarheidsdrempel, kan men zich afvragen of de beperking tot rangtoetsen niet tot een aanzienlijk verlies aan onderscheidingsvermogen leidt. Aangetoond zal worden dat het gebruik van rangtoetsen "asymptotisch" geen verlies aan onderscheidingsvermogen behoeft in te houden. Rangtoetsen zullen voor zekere asymptotische toetsingsproblemen "asymptotisch uniform meest onderscheidende" toetsen blijken te zijn.

We zullen de situatie gaan beschouwen dat $X_1,\ldots,X_N$ opgevat worden als de elementen van de N-de rij uit een driehoekig schema $X_{N1},\ldots,X_{NN}$, N = 1,2,... van rijgewijs o.o. stochastische grootheden (vectoren) met continue k-dimensionale verdelingsfuncties $F_{N1},\ldots,F_{NN}$ en kansmaten $P_{N1},\ldots,P_{NN}$ op $(X,A) = (\mathbb{R}^k,\mathcal{B}^k)$ voor een zeker vast natuurlijk getal k.

Verder zal een (eindig) toetsingsprobleem $(\alpha;H,K)$ ingebed worden in een geschikt gekozen asymptotisch toetsingsprobleem $(\alpha;\{H_N\},\{K_N\})$, waarbij H_N en K_N de nulhypothese en alternatieve hypothese zijn die betrekking hebben op de verdeling van $X_{N1},\ldots,X_{NN}$.

Onder het *asymptotische toetsingsprobleem* $(\alpha,\{H_N\},\{K_N\})$ verstaan we het probleem om, voor iedere N, $P^N \in H_N$ tegen $P^N \in K_N$ te toetsen met zodanige onbetrouwbaarheidsdrempel α_N dat $\limsup_{N\to\infty}\alpha_N \leq \alpha$. We noemen α de *asymptotische onbetrouwbaarheidsdrempel* en $\{\psi_N\}$ een *rij toetsen voor dit probleem*, indien

$$\limsup_{N\to\infty} \int \psi_N dP^N \leq \alpha$$

voor iedere $P^N \in H_N$, $P^N = \Pi_{n=1}^N P_{Nn}$, N = 1,2,... . In hoofdstuk 2 zijn we reeds uitvoerig ingegaan op het asymptotische toetsingsprobleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$, waarbij voor iedere N, een enkelvoudige nulhypothese $\theta(N)$

tegen een enkelvoudige alternatieve hypothese $\tau(N)$ wordt getoetst. We kunnen ons enkelvoudige nulhypothesen $\theta(N)$ en enkelvoudige alternatieve hypothesen $\tau(N)$ ingebed denken in samengestelde nulhypothesen H_N en samengestelde alternatieven hypothesen K_N: d.w.z. $P_{\theta(N)} \in H_N$ respectievelijk $P_{\tau(N)} \in K_N$. We schrijven dan dikwijls $\theta(N) \in H_N$ respectievelijk $\tau(N) \in K_N$.

We beginnen nu het onderzoek naar de asymptotische optimaliteit van rangtoetsen met de definitie van een asymptotisch uniform meest onderscheidende rij toetsen.

DEFINITIE 3.3: Zij $\{\phi_N\}$ een rij toetsen voor het asymptotische toetsingsprobleem $(\alpha;\{H_N\},\{K_N\})$ met $0 < \alpha < 1$. Dan noemt men $\{\phi_N\}$ een *asymptotisch uniform meest onderscheidende* (Engels: asymptotically uniformly most powerful, afgekort tot AUMP) *rij toetsen* voor het probleem $(\alpha;\{H_N\},\{K_N\})$, indien $\{\phi_N\}$ een asymptotisch meest onderscheidende rij toetsen (zie definitie 2.2) is voor ieder probleem $(\alpha;\{H_N\},\{\tau(N)\})$, waarbij voor iedere N, $\tau(N) \in K_N$. Hierbij verstaan we onder een rij toetsen $\{\psi_N\}$ voor het probleem $(\alpha;\{H_N\},\{K_N\})$ een rij toetsen met de eigenschap

$$\limsup_{N\to\infty} \sup_{\theta(N)\in H_N} E_{\theta(N)}\psi_N \leq \alpha.$$

Evenals in het eerste deel van dit hoofdstuk het geval was, zullen we ook nu de theorie in het bijzonder verder ontwikkelen voor het toetsen van de hypothese H_0 van isomorie. We beginnen met de definitie van een asymptotisch toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$ waarin het eindige toetsingsprobleem $(\alpha;H_0,K_c)$ op natuurlijke wijze kan worden ingebed.

We introduceren, voor iedere N, de hypothesen

$$H_{0N} = \{(F_{N1},\ldots,F_{NN}) : F_{N1} = \ldots = F_{NN} = F_N \in F_C\}$$

en, voor gegeven continue verdelingsfuncties $G_{\tau(N,1)},\ldots,G_{\tau(N,N)}$, met $\tau(N,1),\ldots,\tau(N,N) \in T$, gegeven regressieconstanten $c_{N1},\ldots,c_{NN}$ (welke niet allen gelijk zijn) en een vast inwendig punt $\theta_0 \in T$,

$$K_{cN} = \left\{(F_{N1},\ldots,F_{NN}) : F_{Nn} = G_{\tau(N,n)},\ \tau(N,n) = \theta_0 + \Delta_N(c_{Nn}-\bar{c}_N),\right.$$
$$\left. n = 1,\ldots,N,\ \Delta_N > 0\right\},$$

waarbij $\bar{c}_N = \frac{1}{N}\sum_{n=1}^{N} c_{Nn}$. Daar Δ_N de positieve getallen doorloopt is K_{cN} een samengesteld alternatief.

Het asymptotische toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$ betreft het toetsen voor iedere N, van de hypothese H_{0N} van isomorie tegen het regressie-alternatief K_{cN}, met zodanige onbetrouwbaarheidsdrempel α_N dat $\limsup_{N\to\infty}\alpha_N \leq \alpha$. We zullen nu gaan aantonen dat asymptotisch uniform meest onderscheidende rijen toetsen voor het asymptotische toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$, onder voorwaarden welke naburigheid van de onderliggende kansmaten impliceren, gevonden kunnen worden onder de rangtoetsen.

Alvorens we overgaan tot de constructie van een rij asymptotisch uniform meest onderscheidende rangtoetsen voor het probleem $(\alpha,\{H_{0N}\},\{K_{cN}\})$ vermelden we nog een tweetal lemma's.

Laten $U_1,U_2,\ldots$ een rij o.o. identiek verdeelde stochastische grootheden zijn, ieder met homogene verdelingsfunctie op (0,1) en laat voor iedere N, R_{Nn} het rangnummer zijn van U_n in de rij $U_1,\ldots,U_N$ $(n=1,\ldots,N)$.

LEMMA 3.11: Zij γ een kwadratisch-sommeerbare functie op (0,1),

$$\int_0^1 \gamma^2(t)dt < \infty ,$$

dan geldt

$$\lim_{N\to\infty} E(a_N(R_{N1},\gamma) - \gamma(U_1))^2 = 0. \tag{3.21}$$

BEWIJS:
We zullen het lemma hier slechts bewijzen voor het speciale geval dat γ λ-b.o. continu is (zie WITTING-NÖLLE [24], p.139-140). Bij toepassingen is vrijwel steeds aan deze aanname voldaan. Zonder deze aanname is het bewijs aanzienlijk gecompliceerder (zie HÁJEK-ŠIDÁK [12], p. 156-157).

We beginnen het bewijs met, voor iedere N, te definiëren

$$\gamma_N(u) = a_N(1+[uN],\gamma), \qquad 0 < u < 1,$$

waarbij [uN] het grootste natuurlijke getal kleiner of gelijk aan uN voorstelt. Hieruit volgt dat $a_N(R_{N1},\gamma) = \gamma_N(\frac{R_{N1}}{N+1})$, N = 1,2,..., zodat aangetoond moet worden dat

(3.22) $\lim_{N\to\infty} E(\gamma_N(\frac{R_{N1}}{N+1}) - \gamma(U_1))^2 = 0.$

Daartoe merken we allereerst op dat voor iedere N

$$\int_0^1 \gamma_N^2(u)du = \int_0^1 (a_N(1+[uN],\gamma))^2 du = \frac{1}{N}\sum_{n=1}^N a_N^2(n,\gamma) =$$

$$= \frac{1}{N}\sum_{n=1}^N (E\gamma(U^{(N,n)}))^2 \le \frac{1}{N}\sum_{n=1}^N E\gamma^2(U^{(N,n)}) =$$

$$= \frac{1}{N}\sum_{n=1}^N E\gamma^2(U_n) = E\gamma^2(U_1) = \int_0^1 \gamma^2(u)du < \infty,$$

waarbij $U^{(N,n)}$ de n-de geordende waarneming uit een steekproef van omvang N uit een homogene verdeling op (0,1) voorstelt. Hieruit kunnen we concluderen dat

(3.23) $\limsup_{N\to\infty} \int_0^1 \gamma_N^2(u)du \le \int_0^1 \gamma^2(u)du < \infty.$

Vervolgens zullen we laten zien dat

(3.24) $\lim_{N\to\infty} \gamma_N(u) = \gamma(u)$

voor λ-bijna alle $u \in (0,1)$. We beginnen met op te merken dat de dichtheid $f_{N,[Nu]+1}$ van $U^{(N,[Nu]+1)}$ monotoon niet-dalend respectievelijk monotoon niet-stijgend is op $(0;[Nu]/(N-1))$ respectievelijk op $([Nu]/(N-1),1)$ en dat $f_{N,[Nu]+1}(t) \to 0$ voor $t \neq u$, $0 < t,u < 1$, als $N \to \infty$. Verder geldt

$$\gamma_N(u) = \int_0^1 \gamma(t)f(t)_{N,[Nu]+1}(t)dt,$$

voor alle $0 < u < 1$. Maar dit betekent dat voor iedere $\varepsilon > 0$ en voor ieder continuïteitspunt u van γ er een voldoend klein getal $\delta = \delta(\varepsilon,u) > 0$ en een natuurlijk getal $N(\varepsilon)$ bestaat, zodat

$$|\gamma_N(u)-\gamma(u)| \le \int_0^1 |\gamma(t)-\gamma(u)|\, f_{N,[Nu]+1}(t)dt \le$$

$$\le f_{N,[Nu]+1}(u-\delta)\int_0^{u-\delta} |\gamma(t)-\gamma(u)|dt +$$

$$+ f_{N,[Nu]+1}(u+\delta)\int_{u+\delta}^1 |\gamma(t)-\gamma(u)|dt +$$

$$+ \sup_{|u-t|<\delta} |\gamma(u)-\gamma(t)| \int_{u-\delta}^{u+\delta} f_{N,[Nu]+1}(t)dt \le$$

$$\le (f_{N,[Nu]+1}(u-\delta)+f_{N,[Nu]+1}(u+\delta)) \int_0^1 |\gamma(t)-\gamma(u)|dt +$$

$$+ \sup_{|u-t|<\delta} |\gamma(u)-\gamma(t)| \le \varepsilon,$$

voor alle $N \ge N(\varepsilon)$. Hieruit volgt (3.24) onmiddellijk. Toepassing van lemma 3.6 levert nu, op grond van (3.23) en (3.24), dat

$$\lim_{N\to\infty} \int_0^1 \gamma_N^2(u)du = \int_0^1 \gamma^2(u)du. \tag{3.25}$$

Tenslotte bewijzen we (3.22) en daarmee (3.21). We zullen hierbij gebruik maken van (3.25).

$$E(\gamma_N(\tfrac{R_{N1}}{N+1})-\gamma(U_1))^2 =$$

$$= E\gamma_N^2(\tfrac{R_{N1}}{N+1}) + E\gamma^2(U_1) - 2E\gamma_N(\tfrac{R_{N1}}{N+1})\gamma(U_1) =$$

$$= E\gamma_N^2(\tfrac{R_{N1}}{N+1}) + E\gamma^2(U_1) - \tfrac{2}{N}\textstyle\sum_{n=1}^N E(\gamma_N(\tfrac{R_{N1}}{N+1})\gamma(U_1) \mid R_{N1}=n) =$$

$$= \tfrac{1}{N}\textstyle\sum_{n=1}^N \gamma_N^2(\tfrac{n}{N+1}) + \int_0^1 \gamma^2(u)du - \tfrac{2}{N}\textstyle\sum_{n=1}^N \gamma_N^2(\tfrac{n}{N+1}) =$$

$$= \int_0^1 \gamma^2(u)du - \int_0^1 \gamma_N^2(u)du,$$

waarbij de eerste, tweede en vierde overgang triviaal zijn, terwijl de derde overgang berust op lemma 3.1.(b). Toepassing van (3.25) levert nu onmiddellijk (3.22). Hiermee is het bewijs voltooid. □

LEMMA 3.12: Zij N > 1. Laten $(c_1,\ldots,c_N)$ en $(a(1),\ldots,a(N))$ twee willekeurige vectoren zijn en zij

$$T = \sum_{n=1}^{N} c_n a(R_n),$$

waarbij $R_1,\ldots,R_N$ de rangnummers behorende bij een steekproef $X_1,\ldots,X_N$ uit een kansverdeling met continue verdelingsfunctie F voorstellen. Dan geldt

$$(3.26) \qquad ET = \bar{a} \sum_{n=1}^{N} c_n$$

en

$$(3.27) \qquad Var\ T = \frac{1}{N-1} \sum_{n=1}^{N} (c_n-\bar{c})^2 \sum_{n=1}^{N} (a(n)-\bar{a})^2$$

met $\bar{c} = \frac{1}{N} \sum_{n=1}^{N} c_n$ en $\bar{a} = \frac{1}{N} \sum_{n=1}^{N} a(n)$.

BEWIJS: De formules (3.26) en (3.27) volgen vrijwel onmiddellijk uit de definities van verwachting en variantie van een stochastische grootheid. Zie HÁJEK - SIDÁK [12], p.61. □

In het bewijs van de volgende stelling, waarin een asymptotische uniform meest onderscheidende rij rangtoetsen voor het probleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$ wordt gegeven, zullen we essentieel gebruik maken van stelling 2.5 uit hoofdstuk 2. Om de formulering van de stelling te vergemakkelijken herhalen we voorwaarde (I) uit hoofdstuk 2 hier in een vorm die nauw aansluit bij die van voorwaarde (II) uit dit hoofdstuk.

VOORWAARDE (I): Zij θ_0 een inwendig punt van T. De kansmaten behorende bij de familie $\{G_\zeta,\ \zeta \in T\}$ van k-dimensionale verdelingsfuncties worden gedomineerd door een σ-eindige maat μ op $(R^k,\mathcal{B}^k)$. De dichtheid van de kansmaat behorende bij G_ζ ten opzichte van μ noteren we met g_ζ. Er is een verzameling $N_0^c \in \mathcal{B}^k$ met $\mu(N_0) = 0$, zodat (in de eerder vermelde notatie) $g_0(x) > 0$

en $\dot{g}_0(x)$ bestaat voor alle $x \in N_0^c$. Tenslotte geldt

$$(3.28) \qquad \int \dot{g}_0 d\mu = 0, \quad 0 < \int \dot{g}_0^2/g_0 \, d\mu = J_0 < \infty$$

en

$$(3.29) \qquad \lim_{\zeta\to\theta_0} \int (\frac{g_\zeta^{\frac{1}{2}} - g_0^{\frac{1}{2}}}{\zeta - \theta_0})^2 d\mu = \int \lim_{\zeta\to\theta_0} (\frac{g_\zeta^{\frac{1}{2}} - g_0^{\frac{1}{2}}}{\zeta - \theta_0})^2 d\mu \ (= \frac{J_0}{4}).$$

De grootheid J_0, die in (3.29) optreedt, is de *Fisher-informatie* in θ_0 van de familie $\{g_\zeta, \zeta \in T\}$.

We zullen in het vervolg, voor iedere N, de rangnummers van X_{N1} $X_{N1},\ldots,X_{NN}$ met $R_{N1},\ldots,R_{NN}$ aangeven. In het bijzondere geval dat $F_{N1} = \ldots$ $\ldots = F_{NN} = F \in F_C$, waarbij F_{Nn} de verdelingsfunctie van X_{Nn} voorstelt, hebben we in feite te maken met een rij $X_1,X_2,\ldots$ van stochastische grootheden, ieder met continue verdelingsfunctie F. We merken nog op dat, indien $U_n = F(X_n)$, de stochastische grootheden $U_1,U_2,\ldots$ homogeen verdeeld zijn op (0,1), terwijl R_{Nn} dan ook geïnterpreteerd kan worden als het rangnummer van U_n in de rij $U_1,\ldots,U_N$.

STELLING 3.13: Zij k = 1 en θ_0 een inwendig punt van T. Laat de familie $\{G_\zeta, \zeta \in T\}$ aan voorwaarde (I) voldoen en zij $\gamma_0 = \dot{g}_0(G_0^{-1})/g_0(G_0^{-1})$. Voorts voldoen de parameters $c_{N1},\ldots,c_{NN}$ en Δ_N uit K_{cN} aan

$$(3.30) \qquad \Delta_N^2 \max_{1\le n\le N} (c_{Nn}-\bar{c}_N)^2 \to 0, \quad \Delta_N^2 \sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2 \to \eta^2,$$

voor zekere $0 < \eta < \infty$, als $N \to \infty$.
Zij tenslotte

$$(3.31) \qquad \tilde{T}_N = J_0^{-\frac{1}{2}} (\sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2)^{-\frac{1}{2}} \sum_{n=1}^N (c_{Nn}-\bar{c}_N) a_N(R_{Nn},\gamma_0)$$

dan geldt, als $N \to \infty$,

$$(3.32) \qquad L(\tilde{T}_N | P^N) \longrightarrow_z N(0,1)$$

voor iedere $P^N \in H_{0N}$, N = 1,2,... .

Kies $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N \to \infty$, voor zekere $0 < \alpha < 1$. Dan is de rij rangtoetsen $\{\phi_N\} = \{\chi(\{\tilde{T}_N > \tilde{c}_N\})$ asymptotisch uniform meest onderscheidend voor het asymptotische toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$.

Het asymptotische onderscheidingsvermogen is gelijk aan

$$(3.33) \qquad \lim_{N\to\infty} \int \phi_N dP_{\tau(N)} = 1 - \Phi(\xi_{1-\alpha} - \eta J_0^{\frac{1}{2}})$$

voor iedere $\tau(N) \in K_{cN}$, $N = 1,2,\ldots$, waarvoor (3.30) geldt.

BEWIJS:
We beginnen het bewijs met op te merken dat we, voor iedere N, de rangtoetsingsgrootheden $\tilde{T}_N$ zullen gaan vergelijken met stochastische grootheden $\tilde{S}_N$ gedefinieerd door

$$(3.34) \qquad \tilde{S}_N = J_0^{-\frac{1}{2}} \left(\sum_{n=1}^N (\tau(N,n)-\theta_0)^2\right)^{-\frac{1}{2}} \sum_{n=1}^N (\tau(N,n)-\theta_0) \frac{\dot{g}_0(X_{Nn})}{g_0(X_{Nn})} ,$$

waarbij

$$(3.35) \qquad \tau(N,n) = \theta_0 + \Delta_N(c_{Nn} - \bar{c}_N), \qquad n = 1,\ldots,N$$

voor zekere $\Delta_N > 0$. Uit stelling 2.5 volgt nu, op grond van voorwaarde (I), (3.30) en het feit dat (3.30) en (3.35) tezamen relatie (2.20) van stelling 2.5 impliceren, dat

$$(3.36) \qquad L(\tilde{S}_N | P_{\theta(N)}) \longrightarrow_z N(0;1)$$

en

$$(3.37) \qquad L(\tilde{S}_N | P_{\tau(N)}) \longrightarrow_z N(\eta J_0^{\frac{1}{2}};1)$$

voor $\theta(N,n) = \theta_0$ en $\tau(N,n) = \theta_0 + \Delta_N(c_{Nn} - \bar{c}_N)$, $N = 1,2,\ldots$, $n = 1,\ldots,N$. Verder is de rij toetsen $\{\phi_N\} = \{\chi(\{\tilde{S}_N > \tilde{c}_N\})\}$, met $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$ als $N \to \infty$, voor zekere $0 < \alpha < 1$, asymptotisch meest onderscheidend voor het asymptotische toetsingsprobleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ met $\theta(N,n) = \theta_0$ en $\tau(N,n) = \theta_0 + \Delta_N(c_{Nn} - \bar{c}_N)$ voor $N = 1,2,\ldots$, $n = 1,\ldots,N$.

We zullen nu gaan aantonen dat $\tilde{T}_N - \tilde{S}_N$ in $P_{\theta(N)}$-waarschijnlijkheid naar 0 convergeert, als $N \to \infty$. Daartoe merken we allereerst op dat, omdat $\theta(N,n) = \theta_0$, $N = 1,2,\ldots$, $n = 1,\ldots,N$, de o.o. stochastische grootheden $X_{N1},\ldots,X_{NN}$, $N = 1,2,\ldots$ onder de hypothese $\{\theta(N)\}$ identiek verdeeld zijn, ieder met continue verdelingsfunctie G_0. Dit betekent dat we zonder bezwaar de (eerste) index N uit X_{Nn} kunnen weglaten. Verder zullen we dan ook R_n schrijven in plaats van R_{Nn}. Nu impliceert lemma 3.1 dat $(R_1,\ldots,R_N)$ onafhankelijk is van $(X^{(1)},\ldots,X^{(N)})$, de geordende steekproef behorende bij $X_1,\ldots,X_N$. Maar dan is $(R_1,\ldots,R_N)$ ook onafhankelijk van $(U^{(1)},\ldots,U^{(N)})$, waarbij $U_n = G_0(X_n)$, zodat $(U^{(1)},\ldots,U^{(N)})$ een geordende steekproef van omvang N uit een homogene verdeling op (0,1) voorstelt.

Uit het voorafgaande, (3.31), (3.34) en (3.35), volgt nu onmiddellijk dat

$$E_{\theta(N)}((\tilde{T}_N-\tilde{S}_N)^2 \mid U^{(1)}=u^{(1)},\ldots,U^{(N)}=u^{(N)}) =$$

$$= J_0^{-1}(\textstyle\sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2)^{-1}\, E_{\theta(N)}(\sum_{n=1}^N (c_{Nn}-\bar{c}_N)(a_N(R_n,\gamma_0) - \gamma_0(u^{(R_n)})))^2.$$

Vervolgens merken we op dat wegens lemma 3.12

$$E_{\theta(N)}(\textstyle\sum_{n=1}^N (c_{Nn}-\bar{c}_N)(a_N(R_n,\gamma_0) - \gamma_0(u^{(R_n)})))^2 =$$

$$= \frac{1}{N-1} \textstyle\sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2 \sum_{n=1}^N (a_N(n,\gamma_0) - \gamma_0(u^{(n)}) - \bar{a}_N + \bar{\gamma}_{0N})^2$$

met $\bar{a}_N = \frac{1}{N}\sum_{n=1}^N a_N(n,\gamma_0)$ en $\bar{\gamma}_{0N} = \frac{1}{N}\sum_{n=1}^N \gamma_0(u^{(n)})$, zodat

$$E_{\theta(N)}(\textstyle\sum_{n=1}^N (c_{Nn}-\bar{c}_N)(a_N(R_n,\gamma_0) - \gamma_0(u^{(R_n)})))^2 \leq$$

$$\leq \frac{1}{N-1} \textstyle\sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2 \sum_{n=1}^N (a_N(n,\gamma_0) - \gamma_0(u^{(n)}))^2 =$$

$$= \frac{N}{N-1} \textstyle\sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2\, E_{\theta(N)}((a_N(R_1,\gamma_0) - \gamma_0(U_1))^2 | U^{(1)}=u^{(1)},\ldots,U^{(N)}=u^{(N)}).$$

Hieruit volgt onmiddellijk dat

$$E_{\theta(N)}((\tilde{T}_N-\tilde{S}_N)^2) \leq \frac{N}{N-1} J_0^{-1} E_{\theta(N)}(a_N(R_1,\gamma_0) - \gamma_0(U_1))^2.$$

Toepassing van lemma 3.11 (γ_0 is kwadratisch sommeerbaar op grond van de tweede relatie van (3.28)) levert nu dat

$$\lim_{N\to\infty} E_\theta(\tilde{T}_N-\tilde{S}_N)^2 = 0$$

zodat we kunnen concluderen dat

$$(3.38) \qquad \tilde{T}_N - \tilde{S}_N \longrightarrow_{P_{\theta(N)}} 0, \qquad \text{als } N \to \infty.$$

Vervolgens zullen we laten zien dat $\tilde{T}_N - \tilde{S}_N$ ook in $P_{\tau(N)}$-waarschijnlijkheid naar 0 convergeert, als $N \to \infty$. Dit volgt wegens stelling 1.9 onmiddellijk uit (3.38) en het feit dat stelling 2.5, op grond van voorwaarde (I), (3.30) en (3.35) (zie (2.20)), naburigheid van $\{P_{\tau(N)}\}$ t.o.v. $\{P_{\theta(N)}\}$ impliceert. We kunnen concluderen dat

$$(3.39) \qquad \tilde{T}_N - \tilde{S}_N \longrightarrow_{P_{\tau(N)}} 0, \qquad \text{als } N \to \infty.$$

Nu volgt uit (3.36), (3.37), (3.38), (3.39) en lemma 1.8, als $N \to \infty$

$$(3.40) \qquad L(\tilde{T}_N | P_{\theta(N)}) \longrightarrow_z N(0;1)$$

en

$$(3.41) \qquad L(\tilde{T}_N | P_{\tau(N)}) \longrightarrow_z N(\eta J_0^{\frac{1}{2}};1).$$

Verder is de rij rangtoetsen $\{\phi_N\} = \{\chi(\{\tilde{T}_N > c_N\})\}$, met $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$ als $N \to \infty$ voor zekere $0 < \alpha < 1$, asymptotisch meest onderscheidend voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ met $\theta(N,n) = \theta_0$ en $\tau(N,n) = \theta_0 + \Delta_N(c_{Nn}-\bar{c}_N)$, $N = 1,2,\ldots$, $n = 1,\ldots,N$, op grond van stelling 2.5, (3.36), (3.37), (3.40) en (3.41). De uitdrukking (3.33) voor het asymptotische onderscheidingsvermogen is een direct gevolg van (3.40) en (3.41).

Tenslotte volgt (3.32) uit (3.40) en het feit dat iedere rangtoets verdelingsvrij is. Combinatie van dit resultaat met het voorafgaande im-

pliceert, op grond van definitie 3.3, dat de rij rangtoetsen $\{\phi_N\} =$ $= \{\chi(\{\tilde{T}_N > \tilde{c}_N\})\}$, onder de voorwaarden van de stelling, asymptotische uniform meest onderscheidend is voor het probleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$. Hiermee is het bewijs voltooid. □

OPMERKING 3.8: De toetsingsgrootheid $\tilde{T}_N$, welke in (3.31) gedefinieerd wordt, kan ook geschreven worden als

$$(3.42) \qquad \tilde{T}_N = J_0^{-\frac{1}{2}}\left(\sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)^2\right)^{-\frac{1}{2}} \sum_{n=1}^{N} c_{Nn}a_N(R_{Nn},\gamma_0).$$

Om dit in te zien merken we op dat

$$\sum_{n=1}^{N} a_N(n,\gamma_0) = \sum_{n=1}^{N} E\gamma_0(U_n) = N\int_0^1 \gamma_0(t)dt = N\int \dot{g}_0 d\mu = 0,$$

waarbij de laatste overgang volgt uit de eerste relatie in (3.28).

In stelling 3.13 hebben we aangetoond dat de rangtoetsen die op grond van stelling 3.7 lokaal meest onderscheidende rangtoetsen zijn voor het eindige probleem $(\alpha;H_0,K_c)$, onder zekere voorwaarden, asymptotisch optimaal zijn voor het asymptotische probleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$.

We zullen nu met behulp van stelling 3.13 een tweetal resultaten formuleren met betrekking tot de constructie van asymptotisch uniform meest onderscheidende rijen rangtoetsen voor het twee-steekproeven probleem, waarbij H_0 getoetst wordt tegen verschuivingsalternatieven respectievelijk tegen Lehmann-alternatieven van de gedaante (3.18). We hebben dan te maken met het geval dat $c_{Nn} = 1$, $n = 1,\ldots,N_1$, $c_{Nn} = 0$, $n = N_1+1,\ldots,N_1+N_2$, $N_1+N_2 = N$, $N = 1,2,\ldots$. Voor $G_0(G_{\theta_0})$ schrijven we G, terwijl $\Delta_{Nn} = \Delta_N \frac{N_2}{N}$, $n=1,\ldots,N_1$, $\Delta_{Nn} = -\Delta_N \frac{N_1}{N}$, $n = N_1+1,\ldots,N_1+N_2$.

GEVOLG 3.14: Zij G een verdelingsfunctie met een positieve continu-differentieerbare λ-dichtheid g (met afgeleide g'), waarvoor $\int g'^2/g \, d\lambda < \infty$, en zij voor $N = 1,2,\ldots$, $n = 1,\ldots,N$,

(3.43) $$G_{\Delta_{Nn}}(x) = G(x-\Delta_{Nn}), \qquad -\infty < x < \infty,$$

met

(3.44) $$\Delta_N = \eta\sqrt{\frac{N}{N_1N_2}} + \sigma(\sqrt{\frac{N}{N_1N_2}})$$

zodat

$$\begin{cases} \Delta_{Nn} = \eta N_1^{-\frac{1}{2}}N_2^{\frac{1}{2}}N^{-\frac{1}{2}} + o(N_1^{-\frac{1}{2}}) & , n = 1,\ldots,N_1 \\ \quad = -\eta N_1^{\frac{1}{2}}N_2^{-\frac{1}{2}}N^{-\frac{1}{2}} + o(N_2^{-\frac{1}{2}}) & , n = N_1+1,\ldots, N_1+N_2, \end{cases}$$

$N = N_1+N_2$, voor zekere $\eta > 0$. Dan wordt, als $N,N_1,N_2 \to \infty$, een asymptotisch uniform meest onderscheidende rij rangtoetsen voor het twee-steekproeven probleem, waarbij $\{H_{0N}\}$ getoetst wordt tegen rijen verschuivingsalternatieven (3.43) (met (3.44)), gegeven door $\{\phi_N\} = \{\chi(\{\tilde{T}_N > \tilde{c}_N\})\}$, met $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N,N_1,N_2 \to \infty$, voor zekere $0 < \alpha < 1$, en met

(3.45) $$\tilde{T}_N = J_0^{-\frac{1}{2}}N^{-\frac{1}{2}}N_1^{\frac{1}{2}}N_2^{\frac{1}{2}}(N_1^{-1}\sum_{n=1}^{N_1} a_N(R_{Nn},\gamma_0) - N_2^{-1}\sum_{n=N_1+1}^{N} a_N(R_{Nn},\gamma_0))$$

voor iedere N. Verder gelden (3.32) en (3.33).

BEWIJS:
Het bewijs volgt uit stelling 3.13 en het feit dat de voorwaarden van het gevolg die van stelling 3.13 impliceren. We beginnen met op te merken dat de keuze (3.44) aan voorwaarde (3.30) van stelling 3.13 voldoet. Verder volgt onmiddellijk uit het feit dat $c_{Nn} = 1$, $n = 1,\ldots,N_1$, $c_{Nn} = 0$, $n = N_1+1,\ldots,N_1+N_2$, $N_1+N_2 = N$, $N = 1,2,\ldots$ dat $\tilde{T}_N$ de gedaante (3.45) heeft.

Vervolgens tonen we aan dat de voorwaarde opgelegd aan de verdelingsfunctie G voorwaarde (I) met $k = 1$, $N_0^c = \mathbb{R}$ en $\mu = \lambda$ impliceert. Om dit in te zien beginnen we met op te merken dat $\dot{g} = -g'$. Verder geldt

$$\frac{g^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x-\Delta)}{\Delta g^{\frac{1}{2}}(x)} \to \frac{g'(x)}{2g(x)}$$

voor alle $-\infty < x < \infty$ en $\Delta \to 0$. Wegens de ongelijkheid van Schwarz hebben we nu voor iedere $\Delta > 0$

$$0 \leq \left(\frac{g^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x-\Delta)}{\Delta}\right)^2 = \frac{1}{\Delta^2}\left(\int_{x-\Delta}^{x} \frac{g'}{2g^{\frac{1}{2}}}\, d\lambda\right)^2 \leq$$

$$\leq \frac{1}{\Delta}\int_{x-\Delta}^{x} \frac{g'^2}{4g}\, d\lambda = \frac{1}{\Delta}\int \chi_{[x-\Delta,x]}(y)\, \frac{g'^2(y)}{4g(y)}\, dy,$$

waarbij we gebruik maken van de continuïteit van $g'/2g^{\frac{1}{2}}$.

Toepassing van het voorafgaande en van de stelling van Fubini (de integrand is niet negatief) levert nu

$$\int \left(\frac{g^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x-\Delta)}{\Delta g^{\frac{1}{2}}(x)}\right)^2 g(x)dx = \int \left(\frac{g^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x-\Delta)}{\Delta}\right)^2 dx \leq$$

$$\leq \frac{1}{4\Delta}\iint \chi_{[x-\Delta,x]}(y)\, \frac{g'^2(y)}{g(y)}\, dydx =$$

$$= \frac{1}{4\Delta}\iint \chi_{[y,y+\Delta]}(x)dx\, \frac{g'^2(y)}{g(y)}\, dy =$$

$$= \frac{1}{4}\int g'^2/g\, d\lambda = \frac{J_0}{4} < \infty.$$

Uit lemma 3.6 volgt nu (3.29). Tenslotte merken we nog op dat g' sommeerbaar is, omdat

$$\int |g'|\, d\lambda = \int g^{\frac{1}{2}}|g'|/g^{\frac{1}{2}} d\lambda \leq \left(\int g'^2/g d\lambda\right)^{\frac{1}{2}} = J_0^{\frac{1}{2}} < \infty$$

geldt wegens de ongelijkheid van Schwarz. Omdat g' continu is en g een λ-dichtheid voorstelt, levert toepassing van de gedomineerde convergentiestelling

$$\int g'\, d\lambda = \lim_{\substack{M_1 \to -\infty \\ M_2 \to \infty}} \int_{M_1}^{M_2} g'\, d\lambda = \lim_{M_2 \to \infty} g(M_2) - \lim_{M_1 \to -\infty} g(M_1) = 0.$$

Hiermee is het bewijs voltooid. □

VOORBEELD 3.3: De rij rangtoetsen $\{\phi_N\} = \{\chi(\{\tilde{T}_N > \tilde{c}_N\})\}$, met $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N_1, N_2, N \to \infty$, voor zekere $0 < \alpha < 1$, en met $\{\tilde{T}_N\}$ gedefinieerd door

$$\tilde{T}_N = N^{\frac{1}{2}} N_1^{-\frac{1}{2}} N_2^{-\frac{1}{2}} \sum_{n=1}^{N_1} E_V V^{(R_{Nn})}, \qquad N = 1,2,\ldots,$$

waarbij $R_{N1},\ldots,R_{NN_1}$ de rangnummers van $X_{N1},\ldots,X_{NN_1}$ (de "eerste" steekproef) onder $X_{N1},\ldots,X_{NN}$ voorstellen, en $V^{(N1)},\ldots,V^{(NN)}$ een geordende steekproef van omvang N uit de standaard normale verdeling is, is asymptotisch uniform meest onderscheidend voor het toetsen van $\{H_{0N}\}$ tegen rijen normale verschuivingsalternatieven welke voldoen aan (3.43) (met G normaal) en (3.44) voor zekere $\eta > 0$. Dit betekent dat de toets van Fisher-Yates asymptotisch optimaal is voor het twee-steekproeven probleem, waarbij H_0 getoetst wordt tegen normale verschuivingsalternatieven.

Om dit in te zien merken we op dat, indien we voor G de standaard normale verdelingsfunctie Φ kiezen, men eenvoudig verifieert dat $\gamma_0 = \Phi^{-1}$ en $J_0 = 1$. Combinatie van gevolge 3.14 met opmerking 3.8 levert nu onmiddellijk het gewenste resultaat.

VOORBEELD 3.4: De rij rangtoetsen $\{\phi_N\} = \{\chi(\{\tilde{T}_N > \tilde{c}_N\})\}$, met $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N_1, N_2, N \to \infty$, voor zekere $0 < \alpha < 1$, en met $\{\tilde{T}_N\}$ gedefinieerd door

$$T_N = 3^{\frac{1}{2}} N^{\frac{1}{2}} N_1^{-\frac{1}{2}} N_2^{-\frac{1}{2}} \sum_{n=1}^{N_1} \left(2 \frac{R_{Nn}}{N+1} - 1\right), \qquad N = 1,2,\ldots,$$

waarbij $R_{N1},\ldots,R_{NN_1}$ de rangnummers van $X_{N1},\ldots,X_{NN_1}$ onder $X_{N1},\ldots,X_{NN}$ voorstellen, is asymptotisch uniform meest onderscheidend voor het toetsen van $\{H_{0N}\}$ tegen rijen logistieke verschuivingsalternatieven welke voldoen aan (3.43) (met G is logistiek) en (3.44) voor zekere $\eta > 0$. Dit betekent dat de twee-steekproeven toets van Wilcoxon asymptotisch optimaal is voor het twee-steekproeven probleem, waarbij H_0 getoetst wordt tegen logistieke verschuivingsalternatieven.

Om dit in te zien merken we op dat, indien we voor G de (standaard) logistieke verdelingsfunctie (zie voorbeeld 3.2) kiezen, men eenvoudig

verifieert dat $\gamma_0(t) = 2t-1$, $0 < t < 1$. Hieruit volgt direct dat $J_0 = \frac{1}{3}$. Combinatie van gevolg 3.14 met opmerking 3.8 levert onmiddellijk het gewenste resultaat.

Het volgende resultaat betreft de constructie van een asymptotisch uniform meest onderscheidende rij rangtoetsen voor het twee-steekproeven probleem, waarbij H_0 getoetst wordt tegen Lehmann-alternatieven van de gedaante (3.18).

GEVOLG 3.15: Zij G een continue verdelingsfunctie en zij voor $N = 1,2,\ldots$, $n = 1,\ldots,N$

$$G_{\Delta_{Nn}}(x) = (1-\Delta_{Nn})G(x) + \Delta_{Nn}G^2(x), \qquad -\infty < x < \infty, \tag{3.46}$$

met

$$\begin{cases} \Delta_{Nn} = \eta N_1^{-\frac{1}{2}} N_2^{\frac{1}{2}} N^{-\frac{1}{2}} + o(N_1^{-\frac{1}{2}}) & n = 1,\ldots,N_1 \\ \quad = -\eta N_1^{\frac{1}{2}} N_2^{-\frac{1}{2}} N^{-\frac{1}{2}} + o(N_2^{-\frac{1}{2}}) & n = N_1+1,\ldots,N_1+N_2, \end{cases} \tag{3.47}$$

$N = N_1+N_2$, voor zekere $0 < \eta < 1$. Dan wordt, als $N,N_1,N_2 \to \infty$, een asymptotisch uniform meest onderscheidende rij rangtoetsen voor het twee-steekproeven probleem, waarbij $\{H_{0N}\}$ getoetst wordt tegen rijen Lehmann-alternatieven (3.46) met (3.47), voor iedere keuze van $G \in F_C$, gegeven door $\{\phi_N\} = \{\chi(\{\tilde{T}_N > \tilde{c}_N\})\}$, met $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N,N_1,N_2 \to \infty$, voor zekere $0 < \alpha < 1$, en met

$$\tilde{T}_N = 3^{\frac{1}{2}} N^{\frac{1}{2}} N_1^{-\frac{1}{2}} N_2^{-\frac{1}{2}} \sum_{n=1}^{N_1} (2\frac{R_{Nn}}{N+1} - 1), \qquad N = 1,2,\ldots . \tag{3.48}$$

Verder gelden (3.32) en (3.33).

BEWIJS:
Het gevolg is bewezen indien we laten zien dat de voorwaarden van het gevolg die van stelling 3.13 impliceren. Zoals we al in het bewijs van gevolg 3.14 hebben nagegaan volgt voorwaarde (3.30) van stelling 3.13 uit het gegeven (3.47). Verder impliceert opmerking 3.8, de keuze $c_{Nn} = 1$, $n = 1,\ldots,N_1$,

$c_{Nn} = 0$, $n = N_1+1,\ldots,N_1+N_2$, $N_1+N_2 = N$, $N = 1,2,\ldots$ en het feit dat de kansmaten behorende bij verdelingsfuncties G_Δ, $0 < \Delta < 1$, t.o.v. de kansmaat P (dwz. de kansmaat behorende bij de verdelingsfunctie G) P-dichtheden (zie (3.19)) $g_\Delta = 1 + \Delta(2G-1)$ (met afgeleiden $\dot{g}_\Delta = 2G-1$), $0 < \Delta < 1$ bezitten, tezamen met (3.7) dat $\tilde{T}_N$ de gedaante (3.48) heeft.

Vervolgens tonen we aan dat de voorwaarde opgelegd aan de verdelingsfunctie G voorwaarde (I) met $k = 1$, $N_0^c = \mathbb{R}$ en $\mu = P$ impliceert. Uit het voorafgaande volgt dat $\int \frac{\dot{g}^2}{g}\,dP = \int_0^1 (2t-1)^2 dt = \frac{1}{3}$ zodat $J_0 = \frac{1}{3}$. Bovendien geldt dat $\int \dot{g}\,dP = \int_0^1 (2t-1)dt = 0$, waarmee de eerste relatie uit (3.28) is geverifieerd.

Om nu in te zien dat ook (3.29) geldt, merken we allereerst op dat

$$(\tfrac{\partial}{\partial\Delta}\, g_\Delta^{\frac{1}{2}}(x))^2 = \frac{(G(x)-\frac{1}{2})^2}{g_\Delta(x)} \le 2(G(x)-\tfrac{1}{2})^2, \quad -\infty < x < \infty,\ |\Delta| < \tfrac{1}{2},$$

zodat we met behulp van de middelwaarde stelling en de gedomineerde convergentiestelling kunnen laten zien dat

$$\lim_{\Delta\to 0} \int (\frac{g_\Delta^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x)}{\Delta})^2 dP = \int \lim_{\Delta\to 0} (\frac{\partial g_\Delta^{\frac{1}{2}}(x)}{\partial\Delta})^2 dP.$$

Hiermee is het bewijs voltooid. □

In een volgende stelling zullen we laten zien dat onder een extra voorwaarde, stelling 3.13 geldig blijft, indien we de exacte scores $a_N(n,\gamma_0)$, welke in de toetsingsgrootheid (3.31) optreden, vervangen door benaderde scores $a_N(n)$.

STELLING 3.16: Laat aan de voorwaarden van stelling 3.13 voldaan zijn. Zij verder $a_N(n)$, $N = 1,2,\ldots$, $n = 1,\ldots,N$ een driehoekig schema van reële getallen, zodanig dat

$$\lim_{N\to\infty} \int_0^1 (a_N(1+[uN]) - \gamma_0(u))^2 du = 0. \tag{3.49}$$

Zij verder voor iedere N

$$(3.50) \qquad \tilde{T}_N = J_0^{-\frac{1}{2}}\left(\sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)^2\right)^{-\frac{1}{2}} \sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)a_N(R_{Nn}).$$

Dan blijven de resultaten van stelling 3.13 geldig.

BEWIJS:
We beginnen het bewijs met op te merken dat we de rangtoetsingsgrootheid (3.50), met sccres $a_N(n)$, zullen gaan benaderen met de corresponderende rangtoetsingsgrootheid (3.31), met exacte scores $a_N(n,\gamma_0)$.

Op analoge wijze als dat in het betreffende deel van het bewijs van stelling 3.13 is gedaan, kunnen we met behulp van lemma 3.12 laten zien dat

$$E_{\theta(N)}\left(J_0^{-\frac{1}{2}}\left(\sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)^2\right)^{-\frac{1}{2}} \sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)(a_N(R_{Nn})-a_N(R_{Nn},\gamma_0))\right)^2 \leq$$

$$\leq \frac{N}{N-1} J_0^{-1}\int_0^1 (a_N(1+[uN]) - a_N(1+[uN],\gamma_0))^2 du,$$

waarbij $\theta(N,n) = \theta_0$, $N = 1,2,\ldots$, $n = 1,\ldots,N$.

Vervolgens merken we op dat

$$\int_0^1 (a_N(1+[uN]) - a_N(1+[uN],\gamma_0))^2 du \leq$$

$$\leq 2\int_0^1 (a_N(1+[uN]) - \gamma_0(u))^2 du + 2\int_0^1 (a_N(1+[uN],\gamma_0) - \gamma_0(u))^2 du.$$

De eerste term van het rechterlid convergeert op grond van het gegeven (3.49) naar 0, als $N \to \infty$. Om nu in te zien dat ook de tweede term van het rechterlid naar 0 convergeert, als $N \to \infty$, merken we op dat (zie (3.23) en (3.24) met $\gamma = \gamma_0$)

$$\limsup_{N\to\infty} \int_0^1 (a_N(1+[uN],\gamma_0))^2 du \leq \int_0^1 \gamma_0^2(u)du < \infty$$

en

$$\lim_{N\to\infty} a_N(1+[uN],\gamma_0) = \gamma_0(u)$$

voor λ-bijna alle $u \in (0,1)$. Toepassing van lemma 2.4 levert

$$\lim_{N\to\infty} \int_0^1 (a_N(1+[uN],\gamma_0) - \gamma_0(u))^2 du = 0.$$

Uit het voorafgaande volgt nu onmiddellijk dat

$$\lim_{N\to\infty} \int_0^1 (a_N(1+[uN]) - a_N(1+[uN],\gamma_0))^2 du = 0.$$

Maar dit betekent dat

$$\lim_{N\to\infty} E_{\theta(N)}(J_0^{-\frac{1}{2}}(\sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2)^{-\frac{1}{2}} \sum_{n=1}^N (c_{Nn}-\bar{c}_N)(a_N(R_{Nn}) - a_N(R_{Nn},\gamma_0)))^2 = 0,$$

zodat we kunnen concluderen dat

$$(3.51)\quad J_0^{-\frac{1}{2}}(\sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2)^{-\frac{1}{2}} \sum_{n=1}^N (c_{Nn}-\bar{c}_N)(a_N(R_{Nn}) - a_N(R_{Nn},\gamma_0)) \longrightarrow_{P_{\theta(N)}} 0,$$

als $N \to \infty$. Wegens stelling 1.9 en het feit dat de voorwaarden van stelling 3.13 naburigheid van $\{P_{\tau(N)}\}$ ($\tau(N) \in K_{cN}$, voor iedere N) t.o.v. $\{P_{\theta(N)}\}$ impliceert, volgt uit (3.51) dat het linkerlid in (3.51) ook in $P_{\tau(N)}$-waarschijnlijkheid naar 0 convergeert, als $N \to \infty$. Toepassing van lemma 1.8 en stelling 3.13 levert nu onmiddellijk het gewenste resultaat. Hiermee is het bewijs voltooid. □

In het volgende lemma zullen we laten zien dat de benaderde scores $\gamma_0(\frac{n}{N+1})$, onder zekere voorwaarden, aan (3.49) voldoen.

LEMMA 3.17: Zij γ een λ-bijna overal continue kwadratisch-sommeerbare en γ^2 een convexe functie op (0,1). Zij verder

$$(3.52)\qquad a_N(n) = \gamma(\frac{n}{N+1}),$$

voor $N = 1,2,\ldots$, $n = 1,\ldots,N$. Dan geldt (3.49) met $\gamma_0 = \gamma$.

BEWIJS:
We beginnen het bewijs met op te merken dat op grond van het gegeven en de

ongelijkheid van Jensen, voor iedere N,

$$\int_0^1 (a_N(1+[uN]))^2 du = \int_0^1 \gamma^2(\frac{1+[uN]}{N+1})du =$$

$$= \frac{1}{N} \sum_{n=1}^N \gamma^2(\frac{n}{N+1}) = \frac{1}{N} \sum_{n=1}^N \gamma^2(EU^{(N,n)}) \leq$$

$$\leq \frac{1}{N} \sum_{n=1}^N E\gamma^2(U^{(N,n)}) = \frac{1}{N} \sum_{n=1}^N E\gamma^2(U_n) =$$

$$= E\gamma^2(U_1) = \int_0^1 \gamma^2(u)du < \infty,$$

zodat we kunnen concluderen dat

$$\text{limsup}_{N\to\infty} \int_0^1 (a_N(1+[uN]))^2 du \leq \int_0^1 \gamma^2(u)du.$$

Volgens lemma 2.4 is het nu voldoende om te laten zien dat $\lim_{N\to\infty} a_N(1+[uN]) = \gamma(u)$ voor λ-bijna alle $u \in (0,1)$. Maar dit volgt onmiddellijk uit (3.52) en het feit dat λ-bijna overal continu is op $(0,1)$. Hiermee is het bewijs voltooid. □

VOORBEELD 3.5: De rij rangtoetsen $\{\phi_N\} = \{\chi(\{\tilde{T}_N > \tilde{c}_N\})\}$, met $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N_1, N_2, N \to \infty$, voor zekere $0 < \alpha < 1$, en met $\{\tilde{T}_N\}$ gedefinieerd door

$$\tilde{T}_N = N^{\frac{1}{2}} N_1^{-\frac{1}{2}} N_2^{-\frac{1}{2}} \sum_{n=1}^{N_1} \Phi^{-1}(\frac{R_{Nn}}{N+1}), \qquad N = 1,2,\ldots,$$

waarbij $R_{N1},\ldots,R_{NN_1}$ de rangnummers van $X_{N1},\ldots,X_{NN_1}$ onder $X_{N1},\ldots,X_{NN}$ voorstellen, is asymptotisch uniform meest onderscheidend voor het toetsen van $\{H_{0N}\}$ tegen rijen normale verschuivingsalternatieven welke voldoen aan (3.43) (met G normaal) en (3.44) voor zekere $\eta > 0$. Dit betekent dat de toets van Van der Waerden asymptotisch optimaal is voor het twee-steekproevan probleem, waarbij H_0 getoetst wordt tegen normale verschuivingsalternatieven.

Om dit in te zien merken we op dat gevolg 3.14 geldig blijft, indien we de exacte scores, welke in de toetsingsgrootheid (3.45) optreden, vervangen door scores $a_N(n)$ die voldoen aan (3.49). De bewering van dit voorbeeld volgt nu onmiddellijk uit voorbeeld 3.3, lemma 3.17 en het feit dat $(\Phi^{-1})^2$ convex is.

In de tot dusver ontwikkelde asymptotische theorie hebben we ons uitvoerig beziggehouden met de constructie van asymptotisch meest onderscheidende rijen rangtoetsen voor het asymptotische toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$. We hebben gezien dat zekere geschikt gekozen rijen rangtoetsen, welke gebaseerd zijn op rangtoetsingsgrootheden

$$(3.53) \qquad \tilde{T}_N = J_0^{-\frac{1}{2}}\left(\sum_{n=1}^N (c_{Nn}-\bar{c}_N)^2\right)^{-\frac{1}{2}} \sum_{n=1}^N (c_{Nn}-\bar{c}_N)a_N(R_{Nn}),$$

onder zekere voorwaarden, asymptotisch uniform meest onderscheidende rijen toetsen blijken te zijn voor het asymptotsiche toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$. We merken op dat de in (3.53) optredende regressieconstanten $c_{N1},\ldots,c_{NN}$, $N = 1,2,\ldots$, en scores $a_N(1),\ldots,a_N(N)$, $N = 1,2,\ldots$, in zekere zin, optimaal gekozen zijn. De regressieconstanten zijn de, in het alternatief K_{cN} waartegen H_{0N} getoetst wordt, optredende regressieconstanten, terwijl de scores gerelateerd zijn aan de scorefunctie $\gamma_0 = g_0'(G_0^{-1})/g_0(G_0^{-1})$, de scorefunctie behorende bij de familie verdelingsfuncties $\{G_\zeta,\ \zeta \in T\}$ en een inwendig punt $\theta_0 \in T$, welke bij de specificatie van het alternatief K_{cN} gegeven is.

Omdat we in de praktijk dikwijls niet precies weten welk regressiealternatief moet worden getoetst, is het van belang het asymptotische gedrag van rijen rangtoetsen, gebaseerd op rangtoetsingsgrootheden van de gedaante

$$(3.54) \qquad \tilde{T}_N = J^{-\frac{1}{2}}\left(\sum_{n=1}^N (d_{Nn}-\bar{d}_N)^2\right)^{-\frac{1}{2}} \sum_{n=1}^N (d_{Nn}-\bar{d}_N)b_N(R_{Nn}),$$

met willekeurige regressieconstanten $d_{N1},\ldots,d_{NN}$ $(\bar{d}_N = \frac{1}{N}\sum_{n=1}^N d_{Nn})$ en scores $b_N(1),\ldots,b_N(N)$, $N = 1,2,\ldots$, te onderzoeken, voor het geval we deze rij toetsen willen gebruiken voor het probleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$.

Het zal duidelijk zijn dat een rij rangtoetsen, gebaseerd op rangtoetsingsgrootheden (3.54), met willekeurige regressieconstanten $d_{N1},\ldots,d_{NN}$, en scores $b_N(1),\ldots,b_N(N)$, $N = 1,2,\ldots$, in het algemeen niet asymptotisch uniform meest onderscheidend zal zijn voor het asymptotische toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$.

Alvorens we in de volgende stelling hierover een resultaat formuleren, introduceren we eerst nog een voorwaarde voor de regressieconstanten $d_{N1},\ldots,d_{NN}$, $N = 1,2,\ldots$.

NOETHER-VOORWAARDE: Een driehoekig schema van de regressieconstanten $d_{N1},\ldots,d_{NN}$, $N = 1,2,\ldots$, voldoet aan de zgn. Noether-voorwaarde, indien

$$\lim_{N\to\infty} \frac{\max_{1\le n\le N}(d_{Nn}-\bar{d}_N)^2}{\sum_{n=1}^{N}(d_{Nn}-\bar{d}_N)^2} = 0.$$

waarbij $\bar{d}_N = \frac{1}{N}\sum_{n=1}^{N} d_{Nn}$,

STELLING 3.18: Zij $k = 1$ en θ_0 een inwendig punt van T. Laat de familie $\{G_\zeta,\ \zeta \in T\}$ aan voorwaarde (I) voldoen en zij $\gamma_0 = \dot{g}_0(G_0^{-1})/g_0(G_0^{-1})$. Voorts voldoen de parameters $c_{N1},\ldots,c_{NN}$ en Δ_N uit K_{cN} aan

$$(3.55)\qquad \Delta_N^2 \max_{1\le n\le N}(c_{Nn}-\bar{c}_N)^2 \to 0, \quad \Delta_N^2 \sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)^2 \to \eta^2,$$

voor zekere $0 < \eta < \infty$, als $N \to \infty$.

Zij γ een kwadratisch sommeerbare functie op $(0,1)$ en zij

$$(3.56)\qquad \int_0^1 (\gamma(u)-\bar{\gamma})^2 du > 0,$$

waarbij $\bar{\gamma} = \int_0^1 \gamma(u)du$. Zij $b_N(1),\ldots,b_N(N)$, $N = 1,2,\ldots$ een driehoekig schema van scores zódanig dat

$$(3.57)\qquad \lim_{N\to\infty} \int (b_N(1+[uN]) - \gamma(u))^2 du = 0$$

en zij $d_{N1},\ldots,d_{NN}$, N = 1,2,... een driehoekig schema van regressieconstanten welke aan de Noether-voorwaarde voldoet. Bovendien is gegeven dat

$$(3.58)\qquad \lim_{N\to\infty} \frac{\sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)(d_{Nn}-\bar{d}_N)}{(\sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)^2 \sum(d_{Nn}-\bar{d}_N)^2)^{\frac{1}{2}}} = \rho_1 \geq 0,$$

terwijl ρ_2 een niet-negatief getal is, gedefinieerd door

$$(3.59)\qquad \frac{\int_0^1 \gamma(u)\gamma_0(u)du}{(\int_0^1(\gamma(u)-\bar{\gamma})^2du \int_0^1 \gamma_0^2(u)du)^{\frac{1}{2}}} = \rho_2.$$

Zij tenslotte, voor iedere N

$$(3.60)\qquad \tilde{T}_N = J^{-\frac{1}{2}}(\sum_{n=1}^{N}(d_{Nn}-\bar{d}_N)^2)^{-\frac{1}{2}} \sum_{n=1}^{N}(d_{Nn}-\bar{d}_N)b_N(R_{Nn})$$

waarbij $J = \int_0^1(\gamma(u)-\bar{\gamma})^2du$, dan geldt, als $N \to \infty$,

$$(3.61)\qquad L(\tilde{T}_N|P^N) \longrightarrow_z N(0;1)$$

voor iedere $P^N \in H_{0N}$, N = 1,2,... .

Kies $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N \to \infty$, voor zekere $0 < \alpha < 1$. Dan is het asymptotische onderscheidingsvermogen van de rij rangtoetsen $\{\phi_N\} = \{\chi(\{\tilde{T}_N > \tilde{c}_N\})\}$ gelijk aan

$$(3.62)\qquad \lim_{N\to\infty} \int \phi_N dP_{\tau(N)} = 1 - \Phi(\xi_{1-\alpha} - \rho_1\rho_2\eta J_0^{\frac{1}{2}})$$

voor iedere $\tau(N) \in K_{cN}$, N = 1,2,..., waarvoor (3.55) geldt. De asymptotische relatieve doeltreffendheid van de rij toetsen $\{\phi_N\} = \{\chi(\tilde{T}_N > \tilde{c}_N)\}$ met betrekking tot een asymptotisch uniform meest onderscheidende rij toetsen is gelijk aan $\rho^2 = \rho_1^2\rho_2^2$.

BEWIJS:

Het bewijs berust voornamelijk op stelling 2.9 en stelling 3.13. We schrij-

ven $\tilde{T}_N^{bd}$ respectievelijk $\tilde{T}_N^{ac}$ voor (3.60) respectievelijk (3.31). Verder introduceren we, voor iedere N, stochastische grootheden

$$(3.63) \qquad \tilde{S}_N^{bd} = J^{-\frac{1}{2}}\left(\sum_{n=1}^{N}(\lambda(N,n)-\theta_0)^2\right)^{-\frac{1}{2}} \sum_{n=1}^{N}(\lambda(N,n)-\theta_0)\gamma(G_0(X_{Nn})),$$

met

$$\lambda(N,n) = \theta_0 + \Delta_N(d_{Nn}-\bar{d}_N), \qquad n = 1,\ldots,N \text{ en (zie (3.34))}$$

$$(3.64) \qquad \tilde{S}_N^{ac} = J_0^{-\frac{1}{2}}\left(\sum_{n=1}^{N}(\tau(N,n)-\theta_0)^2\right)^{-\frac{1}{2}} \sum_{n=1}^{N}(\tau(N,n)-\theta_0)\frac{\dot{g}_0(X_{Nn})}{g_0(X_{Nn})},$$

met

$$\tau(N,n) = \theta_0 + \Delta_N(c_{Nn}-\bar{c}_N), \qquad n = 1,\ldots,N.$$

We merken op dat de grootheden (3.63) en (3.64) in feite niet van de parameter Δ_N van het alternatief K_{cN} afhangen.

We beginnen het bewijs met op te merken dat we, op grond van voorwaarde (I) en (3.55) reeds hebben laten zien (zie (3.38) en (3.39)) dat $\tilde{T}_N^{ac} - \tilde{S}_N^{ac}$, zowel in $P_{\theta(N)}$ - ($\theta(N,n) = \theta_0$, N = 1,2,..., n = 1,...,N) alsook in $P_{\tau(N)}$-waarschijnlijkheid naar 0 convergeert, als $N \to \infty$. Wegens de ongelijkheid

$$E(\tilde{T}_N^{bd}-\tilde{S}_N^{bd})^2 \le 2\, J^{-1}\left(\sum_{n=1}^{N}(d_{Nn}-\bar{d}_N)^2\right)^{-1} \cdot$$

$$\cdot \left((Eb_N(R_{Nn})-a_N(R_{Nn},\gamma))^2 + E(a_N(R_{Nn},\gamma)-\gamma(U_n))^2\right)$$

is het onder gebruikmaking van (3.56) en (3.57) voldoende argumentatie leidend tot (3.51) respectievelijk (3.38) toe te passen. Verder levert toepassing van stelling 2.5, in combinatie met het feit dat de grootheid $\tilde{S}_N^{ac}$ niet van de parameter Δ_N afhangt, dat de rij toetsen $\{\phi_N\} = \{\chi(\tilde{S}_N^{ac} > \tilde{c}_N)\}$, met $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, voor zekere $0 < \alpha < 1$, als $N \to \infty$, asymptotisch uniform meest onderscheidend is voor het probleem $(\alpha;\{\theta(N)\},\{K_{cN}\})$. Tenslotte volgt uit stelling 2.5 nog (zie (2.22)) dat

$$(3.65) \qquad \eta J_0^{\frac{1}{2}}\tilde{S}_N^{ac} - \Lambda_N \to_{P_{\theta(N)}} \tfrac{1}{2}\eta^2 J_0, \qquad \text{als } N \to \infty,$$

zodat, met $\sigma = \eta J_0^{\frac{1}{2}}$ en $T_{1N} = \tilde{S}_N^{ac}$, aan voorwaarde (2.37) van stelling 2.10. voldaan is. Maar dan is ook aan (3.65) voldaan met $\tilde{T}_N^{ac}$ in plaats van $\tilde{S}_N^{ac}$.

Vervolgens zullen we gaan bewijzen dat de stochastische vector $(\tilde{T}_N^{ac},\tilde{T}_N^{bd})$ aan voorwaarde (2.36) met $\rho = \rho_1\rho_2$ van stelling (2.9) voldoet. Toepassing van stelling 2.10 levert dan de bewering (3.62). Tenslotte geldt dan ook (3.61) omdat voorwaarde (2.36) van stelling 2.10 met $(T_{N1},T_{N2}) = (\tilde{T}_N^{ac},\tilde{T}_N^{bd})$ impliceert dat

$$L(\tilde{T}_N^{bd}|P_{\theta(N)}) \to_z N(0;1).$$

Daar de toets, gebaseerd op $\tilde{T}_N^{bd}$, verdelingsvrij is volgt hieruit onmiddellijk (3.61). We dienen nu nog te bewijzen dat inderdaad

$$(3.66) \qquad L(\tilde{T}_N^{ac},\tilde{T}_N^{bd}|P_{\theta(N)}) \to_z N(0,0;1,1,\rho) \qquad \rho = \rho_1\rho_2$$

geldt. Daar zowel $\tilde{T}_N^{ac} - \tilde{S}_N^{ac}$ als ook $\tilde{T}_N^{bd} - \tilde{S}_N^{bd}$ in $P_{\theta(N)}$-waarschijnlijkheid naar 0 convergeert, als $N \to \infty$, is het op grond van lemma 1.8 en het "Cramer-Wold device" (lemma 1.7) voldoende om te laten zien dat

$$(3.67) \qquad L(\mu_1\tilde{S}_N^{ac} + \mu_2\tilde{S}_N^{bd}|P_{\theta(N)}) \to_z N(0;\mu_1^2+\mu_2^2+2\mu_1\mu_2\rho)$$

voor iedere $\mu_1,\mu_2 \in \mathbb{R}$. Nu is $\mu_1\tilde{S}_N^{ac} + \mu_2\tilde{S}_N^{bd}$ een som van o.o. stochastische grootheden, met termen, welke op grond van voorwaarde (I) en de kwadratische sommeerbaarheid van γ, eindige eerste en tweede momenten bezitten.
Men kan nu op analoge wijze als we in het betreffende deel van het bewijs van stelling 2.5 hebben gedaan, verifiëren dat aan de Lindeberg-voorwaarde van de centrale limietstelling van Lindeberg-Feller is voldaan. Daar waar in het betreffende deel van dat bewijs van stelling 2.5 het gegeven (2.20) gebruikt wordt, dienen we echter voorwaarde (3.30) en de Noether-voorwaarde toe te passen. De bewering over de asymptotische verwachting en variantie in (3.67) volgt direct uit voorwaarde (I), de kwadratische sommeerbaarheid van γ, alsmede (3.58) en (3.59). Hiermee is het bewijs voltooid. □

OPMERKING 3.9: Met behulp van stelling 3.18 kunnen we nu, zonder veel moeite, resultaten formuleren met betrekking tot het asymptotische onderscheidingsvermogen van rijen rangtoetsen voor het twee-steekproeven probleem, onder verschuivingsalternatieven respectievelijk Lehmann-alternatieven.

Tot besluit van dit hoofdstuk vermelden we nog enige resultaten met betrekking tot de asymptotische optimaliteit van de hypothesen H_1 en H_2. We kunnen, voor iedere N, de hypothese H_{1N} en H_{2N} definiëren op een wijze die volkomen analoog is aan de definitie van H_{0N}. Voor wat het toetsen van de hypothese H_1 (symmetrie) betreft beperken we ons hierbij tot de constructie van een rij asymptotisch uniform meest onderscheidende rangtoetsen voor het probleem H_1 te toetsen tegen verschuivingsalternatieven. We schrijven G in plaats van $G_0(G_{\theta_0})$ en we beschouwen verschuivingen over een afstand Δ_N.

STELLING 3.19.: Zij G een symmetrische verdelingsfunctie met symmetrie punt 0 en met een positieve continue-differentieerbare dichtheid g (met afgeleide g'), waarvoor $J_0 = \int g'^2/g d\lambda < \infty$. Zij

$$\gamma_0^+(t) = -\frac{g'(G^{-1}(\frac{1}{2}+\frac{1}{2}t))}{g(G^{-1}(\frac{1}{2}+\frac{1}{2}t))} \qquad 0 < t < 1, \tag{3.68}$$

en zij $a_N^+(n)$, N = 1,2,..., n = 1,...,N een driehoekig schema van reële getallen zodanig dat

$$\lim_{N\to\infty} \int_0^1 (a_N^+(1+[uN]) - \gamma_0^+(u))^2 du = 0. \tag{3.69}$$

Zij verder voor ieder N

$$G_{\Delta_N}(x) = G(x-\Delta_N) \qquad -\infty < x < \infty, \tag{3.70}$$

met

$$\Delta_N = \eta N^{-\frac{1}{2}} + o(N^{-\frac{1}{2}}) \tag{3.71}$$

voor zekere $\eta > 0$.

Zij tenslotte, voor iedere N

$$(3.72) \qquad \tilde{T}_N = J_0^{-\frac{1}{2}} N^{-\frac{1}{2}} \sum_{n=1}^{N} a_N^+(R_{Nn}^+) \operatorname{sgn} X_{Nn}.$$

Dan geldt, als $N \to \infty$,

$$(3.73) \qquad L(\tilde{T}_N | P^N) \longrightarrow_z N(0;1)$$

voor iedere $P^N \in H_{1N}$, $N = 1,2,\ldots$.
Kies $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N \to \infty$, voor zekere $0 < \alpha < 1$. Dan is de rij rangtoetsen $\{\phi_N\} = \{\chi(\{\tilde{T}_N > \tilde{c}_N\})\}$ asymptotisch uniform meest onderscheidend voor het probleem $\{H_{1N}\}$ te toetsen tegen rijen verschuivingsalternatieven (3.70) (met (3.71)).

Het asymptotisch onderscheidingsvermogen is gelijk aan

$$(3.74) \qquad \lim_{N\to\infty} \int \phi_N dP_{\Delta_N} = 1 - \Phi(\xi_{1-\alpha} - \eta J_0^{\frac{1}{2}})$$

voor iedere $\{\Delta_N\}$ waarvoor (3.71) geldt.

BEWIJS:
Het bewijs verloopt analoog aan het bewijs van gevolg 3.14. Omdat we hier te maken hebben met benaderde scores dienen we ook nog gebruik te maken van (3.69) (zie stelling 3.16). □

Tenslotte geven we een resultaat betreffende de asymptotische optimaliteit van de rangcorrelatie toets van Spearman voor het toetsen van de hypothese H_2 van onafhankelijkheid tegen zekere Lehmann-alternatieven.

STELLING 3.20: Zij voor iedere N

$$(3.75) \qquad G_{\Delta_N}(x) = H(y)K(z)(1+\Delta_N(H(y)-1)(K(z)-1)), \qquad x = (y,z),$$

$-\infty < y,z < \infty$, $0 < \Delta < 1$, $H,K \in F_C$, met

$$(3.76) \qquad \Delta_N = \eta N^{-\frac{1}{2}} + o(N^{-\frac{1}{2}})$$

voor zekere $0 < \eta < 1$.

Zij tenslotte, voor iedere N

$$(3.77) \qquad \tilde{T}_N = 3N^{-\frac{1}{2}} \sum_{n=1}^{N} (2\frac{R_{Nn}}{N+1} - 1)(2\frac{S_{Nn}}{N+1} - 1).$$

Dan geldt, als $N \to \infty$,

$$(3.78) \qquad L(\tilde{T}_N | P^N) \longrightarrow_z N(0;1)$$

voor iedere $P^N \in H_{2N}$, $N = 1,2,\ldots$.

Kies $\{\tilde{c}_N\}$ zodanig dat $\tilde{c}_N \to \xi_{1-\alpha}$, als $N \to \infty$, voor zekere $0 < \alpha < 1$. Dan is de rij rangtoetsen $\{\phi_N\} = \{\chi(\tilde{T}_N > \tilde{c}_N)\}$ asymptotisch uniform meest onderscheidend voor het probleem $\{H_{2N}\}$ te toetsen tegen Lehmann-alternatieven (3.75), voor iedere keuze van H en $K \in F_C$.

Het asymptotisch onderscheidingsvermogen is gelijk aan

$$(3.79) \qquad \lim_{N\to\infty} \int \phi_N dP_{\Delta_N} = 1 - \Phi(\xi_{1-\alpha} - 3^{-1}\eta).$$

BEWIJS:
Het bewijs verloopt analoog aan het bewijs van gevolg 3.15. Hierbij vervangen we stelling 3.13 door een analoge stelling, welke hier onvermeld blijft, betreffende de rangtoetsingsgrootheden (3.77). Tenslotte dienen we dan te verifiëren dat de voorwaarden van het gevolg voorwaarde (I) met k = 2, $N_0^c = \mathbb{R}^2$ en $\mu = P^2$ (de kansmaat behorende bij de verdelingsfunctie G = H.K, $H,K \in F_c$) en (2.20) impliceert. □

BRONVERMELDINGEN

De uiteenzetting over de theorie der rangtoetsen is grotendeels ontleend aan HÁJEK & ŠIDÁK [12] en WITTING & NÖLLE [24]. De constructie van lokaal meest onderscheidende rangtoetsen voor het toetsen van de hypothese H_0 van isomorie tegen een regressie-alternatief K_c in stelling 3.7, is een enigszins gewijzigde versie van een resultaat uit HÁJEK & ŠIDÁK [12]. Voorwaarde (II) van stelling 3.7, welke nauw aansluit bij voorwaarde (I) uit hoofdstuk 2, is een variant van de in stelling II.4.8 van HÁJEK - ŠIDÁK [12] vermelde voorwaarde A1. De formulering van gevolg 3.8 ontleend aan HÁJEK - ŠIDÁK [12], is in verband hiermee enigszins gewijzigd.

De behandeling van de asymptotische optimaliteit van rangtoetsen berust grotendeels op HÁJEK - ŠIDÁK [12] en WITTING - NÖLLE [24]; zie ook HÁJEK [10]. De definitie (definitie 3.3) van een asymptotisch uniform meest onderscheidende rij toetsen is overgenomen uit WITTING - NÖLLE [24]. De formulering van stelling 3.13, evenals stelling 3.16 en 3.18, zijn grotendeels ontleend aan HÁJEK - ŠIDÁK [12]. Voorwaarde (I) -zie ook hoofdstuk 2- is overgenomen uit WITTING - NÖLLE [24], terwijl ook voorwaarde (3.30) van stelling 3.13, voor het speciale geval dat $C_{Nn} = 1$, $n = 1,\ldots,N_1$, $C_{Nn} = 0$, $n = N_1+1,\ldots, N_1 + N_2$, $N_1 + N_2 = N$, te vinden is in WITTING - NÖLLE [24]. Gevolg 3.14, 3.15, lemma 3.17, stelling 3.19 en 3.20 zijn eveneens te vinden in WITTING - NÖLLE [24].

Van de meer recente literatuur over asymptotische optimaliteit van rangtoetsen noemen we BERAN [5] en BEHNEN [3], [4]. Een uitstekende inleiding tot de in dit hoofdstuk besproken theorie is te vinden in HÁJEK [11].

Vraagstukken

1.1. Laten voor $N = 1,2,\ldots$, P_N en Q_N kansmaten zijn, gedefinieerd op de meetbare ruimte $(\mathcal{X}_N,\mathcal{A}_N)$.

Laat zien dat de volgende implicaties *niet* juist zijn:

a) $\{P_N\} \Leftrightarrow \{Q_N\} \Rightarrow P_N \approx Q_N$ voor voldoend grote N, (waarin $P \approx Q$ aangeeft dat P en Q equivalent zijn),

b) $P_N \approx Q_N$ voor $N = 1,2,\ldots \Rightarrow \{P_N\} \Leftrightarrow \{Q_N\}$,

c) $\{P_N\} \Leftrightarrow \{Q_N\} \Rightarrow \| P_N - Q_N \| \to 0$ als $N \to \infty$.

In de opgaven 1.2 t/m 1.5 worden een aantal voorbeelden gegeven van situaties waarin twee rijen kansmaten naburig zijn.

1.2. Beschouw voor $\theta > -1$ de familie van kansdichtheden t.o.v. de Lebesguemaat op $(\mathbb{R},\mathcal{B})$:

$$f_\theta(x) = \begin{cases} (\theta+1)\, x^\theta & \text{voor } x \in (0,1), \\ 0 & \text{voor } x \in (0,1)^c. \end{cases}$$

Zij voor $N = 1,2,\ldots, X_N$ een stochastische grootheid met óf P_{θ_N} óf $P_{\theta'_N}$ als kansverdeling op $(\mathcal{X}_N,\mathcal{A}_N) = (\mathbb{R},\mathcal{B})$, bepaald door de dichtheden f_{θ_N} respectievelijk $f_{\theta'_N}$.

Als nu gegeven is dat de rijen $\{\theta_N\}$ en $\{\theta'_N\}$ beide naar boven begrensd en van -1 verwijderd zijn, bewijs dan dat:

$$\{P_{\theta_N}\} \Leftrightarrow \{P_{\theta'_N}\}.$$

1.3. Een klasse $\mathcal{P} = \{P_\theta,\ \theta \in (a,b) \subset \mathbb{R}\}$ van kansmaten op $(\mathbb{R},\mathcal{B})$ heet een een-parameter exponentiële familie in θ en x als er een dominerende

maat ν bestaat met de eigenschap dat de ν-dichtheid f_θ van P_θ van de volgende vorm is:

$$f_\theta(x) = c(\theta)\exp(\theta x).$$

Laten $\{\theta_N\}$ en $\{\theta'_N\}$ twee rijen reële getallen in (a,b) zijn, verwijderd van zowel a als b. Bewijs dat:

$$\{P_{\theta_N}\} \Phi \{P_{\theta'_N}\}.$$

1.4. Laten voor $N = 1,2,\ldots,P_N^N$ en Q_N^N de simultane verdelingen zijn van N onafhankelijke volgens P_{θ_0} respectievelijk P_{θ_N} verdeelde stochastische grootheden $X_{N1},\ldots,X_{NN}$, waarbij P_θ element is van een een-parameter exponentiële familie in θ en $t(x)$, met dichtheid van de vorm

$$f_\theta(x) = c(\theta)\exp(\theta t(x)),\ \theta \in \Theta,$$

t.o.v. een dominerende maat ν. Bewijs dat:

$$\{P_N^N\} \Phi \{Q_N^N\} \quad \text{als } \theta_0 \in \text{int}(\Theta) \text{ en } \theta_N = \theta_0 + \eta N^{-\frac{1}{2}},$$
$$\text{voor zekere } \eta \in \mathbb{R}.$$

1.5. Laten P_N^N en Q_N^N de simultane verdelingen zijn van N onafhankelijke $N(\mu_N,\sigma_N^2)$ respectievelijk $N(\nu_N,\sigma_N^2)$ verdeelde stochastische grootheden $X_{N1},\ldots,X_{NN}$. Bewijs dat geldt:

$$\{P_N^N\} \Phi \{Q_N^N\} \Leftrightarrow N\left(\frac{\mu_N-\nu_N}{\sigma_N}\right)^2 \text{ is begrensd.}$$

Bij rijen o.o.i. verdeelde stochastische grootheden impliceert naburigheid van simultane kansverdelingen dat de marginale kansverdelingen tot elkaar naderen, zoals in opgave 1.7 preciezer wordt geformuleerd. Deze implicatie geeft een goed inzicht in de betekenis van naburigheid van rijen simultane kansverdelingen.

1.6. Laat voor $N = 1,2,\ldots$, P_N en Q_N kansmaten zijn op (X_N,A_N). Dan geldt de implicatie:

$$\{P_N\} \Leftrightarrow \{Q_N\} \Rightarrow \limsup_{N\to\infty} \| P_N - Q_N \| < 1,$$

(vergelijk stelling 1.12).

1.7. Laten voor $N = 1,2,\ldots$, P_N en Q_N kansmaten zijn op de meetbare ruimte (X_N,A_N).
Beschouw voor $N = 1,2,\ldots$, $P_N^N := P_N \times P_N \times \ldots \times P_N$ en $Q_N^N := Q_N \times Q_N \times \ldots \times Q_N$, de N-voudige productmaten, gedefinieerd op de productruimte (X_N^N, A_N^N).

a) Laat zien dat in deze situatie de volgende implicatie juist is:

$$\{P_N^N\} \Leftrightarrow \{Q_N^N\} \Rightarrow \| P_N - Q_N \| \to 0 \quad \text{voor } N\to\infty,$$

zodat bij naburigheid van de productmaten de marginalen in sterke zin tot elkaar naderen als $N\to\infty$.

b) Bewijs dat *niet* geldt:

$$\| P_N - Q_N \| \to 0 \quad \text{voor } N\to\infty \Rightarrow \{P_N^N\} \Leftrightarrow \{Q_N^N\}.$$

1.8. Laten voor $N = 1,2,\ldots$, P_N en Q_N kansmaten zijn, gedefinieerd op de meetbare ruimte (X_N,A_N).
Beschouw voor $N = 1,2,\ldots$, de productmaten $P^N := \prod_{i=1}^N P_i$ en $Q^N := \prod_{i=1}^N Q_i$, gedefinieerd op de productmaatruimte $(\prod_{i=1}^N X_i, \prod_{i=1}^N A_i)$.

a) Bewijs dat de volgende implicatie juist is:

$$\{P^N\} \Leftrightarrow \{Q^N\} \Rightarrow \{P_N\} \Leftrightarrow \{Q_N\} \text{ én } P_N \approx Q_N \quad \text{voor } N = 1,2,\ldots .$$

b) Laat zien dat de omkering van a) niet juist is.

1.9. Laat aan de hand van een tegenvoorbeeld zien dat de condities (b') en (c') van stelling 1.14 niet impliceren $\{P_N\} \Leftrightarrow \{Q_N\}$.

In stelling 1.9 en stelling 1.13 zijn karakterizeringen gegeven van het begrip naburigheid van twee rijen kansmaten. Opgave 1.10 geeft nog een andere karakterisering.

1.10. Laten voor $N = 1,2,\ldots$, P_N en Q_N kansmaten zijn gedefinieerd op de meetbare ruimte (X_N,A_N). Bewijs dat de volgende uitspraken equivalent zijn:

a) $\{P_N\} \Leftrightarrow \{Q_N\}$,

b) voor alle A_N-meetbare stochastische grootheden T_N $(N = 1,2,\ldots)$ geldt:

$\{L(T_N|P_N)\}$ is relatief compact $\Leftrightarrow$ $\{L(T_N|Q_N)\}$ is relatief compact.

1.11. Laten voor $N = 1,2,\ldots$, P_N en Q_N kansmaten zijn gedefinieerd op (X_N,A_N), laat T_N een A_N-meetbare k-dim. stochastische vector zijn, $(k \geq 1)$ en laat Λ_N de logarithme zijn van het aannemelijkheidsquotiënt, zoals gedefinieerd in 1.14.

Als $\{L((\Lambda_N,T_N)|P_N)\} \xrightarrow{z} N(\mu;\Sigma)$, voor zekere (k+1)-dim. verwachtingsvector $\mu = (-\frac{1}{2}\sigma^2,\mu_1,\ldots,\mu_k)$ en

$$\text{covariantiematrix } \Sigma = \begin{pmatrix} \sigma^2 & \sigma_{01} & \sigma_{02} & \cdots & \sigma_{0k} \\ \sigma_{01} & \sigma_{11} & & & \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ \sigma_{0k} & & & & \sigma_{kk} \end{pmatrix},$$

dan geldt $\{L((\Lambda_N,T_N)|Q_N)\} \to N(\nu;\Sigma)$, met $\nu = \mu + \Sigma_1$, waarbij Σ_1 de eerste rijvector is van Σ.

Bewijs dit.

1.12. Laten voor $N = 1,2,\ldots$, P_N en Q_N kansmaten zijn, gedefinieerd op (X_N,A_N). Laat vervolgens voor $N = 1,2,\ldots$, T_N een k-dim. stochastische vector ($k \geq 1$) zijn op (X_N,A_N) en laat Λ_N de logarithme zijn van het aannemelijkheidsquotiënt, zoals gedefinieerd in 1.14.
Stel dat $\mathcal{L}(T_N|P_N) \xrightarrow{z} N(0;\Gamma)$, waar Γ een niet-singuliere $k \times k$-covariantiematrix is. Laat er tenslotte een $h \in \mathbb{R}^k$ bestaan zó dat geldt:

$$hT_N' - \Lambda_N \to_{P_N} \frac{1}{2}\sigma^2,$$

waar $\sigma^2 = \sigma^2(h) := h\Gamma h'$. Bewijs dat $\mathcal{L}((\Lambda_N,T_N)|P_N) \xrightarrow{z} N(\mu;\Sigma)$, waarbij

$$\mu = (-\tfrac{1}{2}\sigma^2,0,\ldots,0) \text{ en } \Sigma = \begin{pmatrix} \sigma^2 & h\Gamma \\ \Gamma h' & \Gamma \end{pmatrix}.$$

Opmerking: Uit opgaven 1.11 en 1.12 leiden we nu direct de volgende generalisatie van gevolg 1.18(iii) af:
Zij T_N een stochastische k-dim. vector ($k \geq 1$) op (X_N,A_N), $N = 1,2,\ldots$ en zij $\sigma^2 = h\Gamma h'$ voor zekere $k \times k$-covariantiematrix Γ en zekere $h \in \mathbb{R}^k$, dan geldt de volgende implicatie:

$$\left.\begin{array}{l} hT_N' - \Lambda_N \to_{P_N} \frac{1}{2}\sigma^2 \\ \mathcal{L}(T_N|P_N) \xrightarrow{z} N(0,\Gamma) \end{array}\right\} \Rightarrow \mathcal{L}((\Lambda_N,T_N)|Q_N) \xrightarrow{z} N(\nu;\Sigma),$$

met $\Sigma = \begin{pmatrix} \sigma^2 & h\Gamma \\ \Gamma h' & \Gamma \end{pmatrix}$ en $\nu = (\frac{1}{2}\sigma^2,h\Gamma)$.

Oplossingen der vraagstukken

1.1. We geven een aantal tegenvoorbeelden:

a) Zij P_N de homogene verdeling op $(0,1)$ voor $N = 1,2,\ldots$, en zij Q_N de homogene verdeling op $(\frac{1}{N}, 1+\frac{1}{N})$ voor $N = 1,2,\ldots$. Dan is $P_N \not\approx Q_N$ (alle N) en $\|P_N - Q_N\| = \frac{1}{N} \to 0$ als $N \to \infty$.

b) Zij P_N de normale $N(\mu_N;1)$ verdeling $(N=1,2,\ldots)$ met $\lim_{N\to\infty} \mu_N = \infty$ en zij Q_N de normale $N(\nu_N;1)$ verdeling $(N=1,2,\ldots)$ met $\limsup_{N\to\infty} \nu_N < \infty$. Kies $A_N = (\mu_N, \infty)$ voor $N = 1,2,\ldots$, dan geldt $\lim_{N\to\infty} P_N(A_N) = \frac{1}{2}$ en $\lim_{N\to\infty} Q_N(A_N) = 0$ zodat $\{P_N\} \not\triangleleft \{Q_N\}$. Anderzijds is $P_N \approx Q_N$ voor alle N.

c) Zij P_N de normale $N(\mu_N;1)$ verdeling $(N=1,2,\ldots)$ met $\lim_{N\to\infty} \mu_N = \mu \in \mathbb{R}$ en zij Q_N de normale $N(\nu_N;1)$ verdeling $(N=1,2,\ldots)$ met $\lim_{N\to\infty} \nu_N = \nu \in \mathbb{R}$, waarin $\mu \neq \nu$. Volgens voorbeeld 1.2 is $\{P_N\} \triangleleft\triangleright \{Q_N\}$, terwijl

$$\lim_{N\to\infty} \|P_N - Q_N\| = \lim_{N\to\infty} |\Phi(\tfrac{1}{2}(\nu_N-\mu_N)) - \Phi(\tfrac{1}{2}(\mu_N-\nu_N))|$$
$$= |2\Phi(\tfrac{1}{2}(\nu-\mu)) - 1| > 0.$$

1.2. De grootheid Λ_N uit formule (1.14) heeft thans met kans 1 de gedaante $\Lambda_N = \Lambda_N(X_N) = \log\{(\theta'_N+1)/(\theta_N+1)\} + (\theta'_N-\theta_N)\log X_N$, $N = 1,2,\ldots$. Elke deelrij $\{m\} \subset \{N\}$ bevat een subdeelrij $\{r\}$ zodat $\theta_r \to \theta$ en $\theta'_r \to \theta'$ als $r \to \infty$ voor zekere $-1 < \theta < \infty$ en $-1 < \theta' < \infty$. Zij P_θ (respectievelijk $P_{\theta'}$) de kansverdeling met dichtheid f_θ (respectievelijk $f_{\theta'}$). Dan geldt $P_r \longrightarrow_z P_\theta$ en $Q_r \longrightarrow_z P_{\theta'}$. Hieruit volgt op grond van de eerste bewering van lemma 1.4 dat

$$\Lambda_r \longrightarrow_D \log\{(\theta'+1)/(\theta+1)\} + (\theta'-\theta)\log X \quad \text{voor } r \to \infty,$$

waarin de kansverdeling van de stochastische grootheid X in het eerste geval P_θ en in het tweede geval $P_{\theta'}$ is. Hiermee is bewezen dat aan eigenschap (b) van stelling 1.13 voldaan is, zodat op grond van deze stelling $\{P_{\theta_N}\} \triangleleft\triangleright \{P_{\theta'_N}\}$.

1.3. De grootheid Λ_N uit formule (1.14) heeft met kans 1 de gedaante $\Lambda_N = \Lambda_N(X_N) = \log(c(\theta_N)/c(\theta_N')) + (\theta_N-\theta_N')\,X_N$, $N = 1,2,\ldots$. Elke deelrij $\{m\} \subset \{N\}$ bevat een subdeelrij $\{r\}$ zodat $\theta_r \to \theta$ en $\theta_r' \to \theta'$ als $r \to \infty$ voor zekere $a < \theta < b$ en $a < \theta' < b$. Zij P_θ (respectievelijk $P_{\theta'}$) de kansverdeling met dichtheid f_θ (respectievelijk $f_{\theta'}$). Dan geldt $P_r \to_z P_\theta$ en $Q_r \to_z P_{\theta'}$.

Het is algemeen bekend dat $c(\theta)$ een continue functie van θ is op (a,b) zodat volgt dat

$$\Lambda_r \to_D \log(c(\theta)/c(\theta')) + (\theta-\theta')\,X \qquad \text{voor } r \to \infty,$$

waarin de kansverdeling van de stochastische grootheid X in het eerste geval P_θ en in het tweede geval $P_{\theta'}$ is. Hiermee is bewezen dat aan eigenschap (b) van stelling 1.13 voldaan is, zodat op grond van deze stelling $\{P_{\theta_N}\} \Phi \{P_{\theta_N'}\}$.

1.4. De grootheid Λ_N uit formule (1.14) heeft thans met kans 1 de gedaante

$$\Lambda_N = \Lambda_N(X_{N1},\ldots,X_{NN}) = N\log\{c(\theta_N)/c(\theta_0)\} + (\theta_N-\theta_0)\sum_{i=1}^{N} t(X_{Ni}) =$$

$$= N\log\{c(\theta_0+\eta N^{-\frac{1}{2}})/c(\theta_0)\} + \eta N^{-\frac{1}{2}}\sum_{i=1}^{N} t(X_{Ni}).$$

Het is algemeen bekend dat $c(\cdot)$ een oneindig vaak differentieerbare functie is op int (Θ). Hieruit volgt dat

$$\log c(\theta_0+\eta N^{-\frac{1}{2}}) = \log c(\theta_0) + \eta N^{-\frac{1}{2}}\frac{c'(\theta_0)}{c(\theta_0)} +$$

$$+ \tfrac{1}{2}\eta^2 N^{-1}\frac{c''(\theta_0)c(\theta_0)-\{c'(\theta_0)\}^2}{\{c(\theta_0)\}^2} + O(N^{-3/2}).$$

Men kan laten zien dat voor alle $i = 1,2,\ldots,N$,

$$E_{\theta_0} t(X_{Ni}) = -\frac{c'(\theta_0)}{c(\theta_0)} \quad \text{en} \quad var_{\theta_0} t(X_{Ni}) = \frac{\{c'(\theta_0)\}^2-c''(\theta_0)c(\theta_0)}{\{c(\theta_0)\}^2}.$$

Hieruit volgt dat

$$\Lambda_N = \tfrac{1}{2}\eta^2 \frac{c''(\theta_0)c(\theta_0)-\{c'(\theta_0)\}^2}{\{c(\theta_0)\}^2} + \eta N^{-\frac{1}{2}} \sum_{i=1}^{N} \{t(X_{Ni})-Et(X_{Ni})\}+o(N^{-\frac{1}{2}}).$$

Op grond van de centrale limietstelling convergeert de tweede term in het linkerlid in verdeling naar de normale

$$N(0;\eta^2 \frac{\{c'(\theta_0)\}^2-c''(\theta_0)c(\theta_0)}{\{c(\theta_0)\}^2})$$

verdeling.

Maar dan geldt ook (zie lemma 1.8) dat

$$L(\Lambda_N|P_N^N) \to_z N(-\tfrac{1}{2}\sigma^2;\sigma^2) \text{ met } \sigma^2 = \eta^2 \frac{\{c(\theta_0)\}^2-c''(\theta_0)c(\theta_0)}{\{c(\theta_0)\}^2}$$

zodat wegens gevolg 1.16 $\{P_N^N\} \lhd\rhd \{Q_N^N\}$.

1.5. We berekenen voor N = 1,2,... de logarithme van het aannemelijkheidsquotient:

$$\Lambda_N = \Lambda_N(X_{N1},\ldots,X_{NN}) = \log\Big[\prod_{i=1}^{N} \frac{(2\pi\sigma_N^2)^{-\frac{1}{2}}\exp\{-\frac{1}{2}\sigma_N^{-2}(X_{Ni}-\mu_N)^2\}}{(2\pi\sigma_N)^{-\frac{1}{2}}\exp\{-\frac{1}{2}\sigma_N^{-2}(X_{Ni}-\nu_N)^2\}}\Big] =$$

$$= \tfrac{1}{2}\sum_{i=1}^{N} \{\sigma_N^{-2}(X_{Ni}-\nu_N)^2 - \sigma_N^{-2}(X_{Ni}-\mu_N)^2\} =$$

$$= \sigma_N^{-2}(\mu_N-\nu_N)\sum_{i=1}^{N} X_{Ni} + \tfrac{1}{2}N\sigma_N^{-2}(\nu_N^2-\mu_N^2).$$

Als $L(X_{Ni}) = N(\mu_N;\sigma_N^2)$ voor i = 1,2,...,N, dan geldt

$$L(\Lambda_N) = N(-\tfrac{1}{2}N(\frac{\mu_N-\nu_N}{\sigma_N})^2;N(\frac{\nu_N-\mu_N}{\sigma_N})^2).$$

Men ziet eenvoudig in dat een rij $N(a_N;\tau_N^2)$ verdelingen d.e.s.d. uniform beperkt is indien de rijen $\{a_N\}$ en $\{\tau_N^2\}$ begrensd zijn. Hieruit volgt op grond van lemma 1.5 dat $\{L(\Lambda_N|P_N^N)\}$ d.e.s.d. relatief compact

is als de rij $\{N(\frac{\mu_N-\nu_N}{\sigma_N})^2\}$ begrensd is. Geheel analoog laat men nu zien dat $\{L(\Lambda_N|Q_N^N)\}$ d.e.s.d. relatief compact is als de rij $\{N(\frac{\mu_N-\nu_N}{\sigma_N})^2\}$ begrensd is. Toepassing van stelling 1.13 geeft nu het verlangde resultaat.

1.6. Onderstel dat $\limsup_{N\to\infty}\|P_N-Q_N\| = 1$. Dan is er een deelrij $\{m\}$ van $\{N\}$ zodat $\lim_{m\to\infty}\|P_m-Q_m\| = 1$. Dit impliceert dat er een rij verzamelingen $\{A_m\}$ bestaat zodat $\lim_{m\to\infty}|P_m(A_m)-Q_m(A_m)| = 1$.
Maar dan bestaat er een deelrij $\{r\}$ van $\{m\}$ met de eigenschap dat

óf $\lim_{r\to\infty} P_r(A_r) = 1$ en $\lim_{r\to\infty} Q_r(A_r) = 0$,

óf $\lim_{r\to\infty} Q_r(A_r) = 1$ en $\lim_{r\to\infty} P_r(A_r) = 0$.

Dit is in strijd met de naburigheid van $\{P_N\}$ en $\{Q_N\}$.

1.7. a) Uit $\{P_N^N\} \not\sim \{Q_N^N\}$ volgt met behulp van opgave 1.6 dat er een $\varepsilon > 0$ bestaat zodat $\delta(P_N^N,Q_N^N)^2 < 1-\varepsilon$ voor alle voldoend grote N.
Zij $\mu_N = P_N + Q_N$, N = 1,2,... . Dan geldt voor alle voldoend grote N

$$2-\varepsilon > \delta(P_N^N,Q_N^N)^2 = \int\cdots\int(\prod_{n=1}^N p_N(x_n)^{\frac12} - \prod_{n=1}^N q_N(x_n)^{\frac12})^2 d\mu_N(x_1)\ldots$$

$$\ldots d\mu_N(x_N) =$$

$$= 2 - 2\int\cdots\int\prod_{n=1}^N p_N(x_n)^{\frac12}q_N(x_n)^{\frac12}d\mu_N(x_1)\ldots d\mu_N(x_N) =$$

$$= 2 - 2\prod_{n=1}^N\int p_N(x_n)^{\frac12}q_N(x_n)^{\frac12}d\mu_N(x_n) = 2 - 2\prod_{n=1}^N\left[1 - \tfrac12\delta(P_N,Q_N)^2\right]$$

$$= 2 - 2\left[1 - \tfrac12\delta(P_N,Q_N)^2\right]^N ,$$

zodat $[1 - \frac12\delta(P_N,Q_N)^2]^N > \frac12(1+\varepsilon)$ voor alle voldoend grote N. Dit impliceert dat $\lim_{N\to\infty}\delta(P_N,Q_N) = 0$.
Overgang op logarithme geeft

$$\log[1 - \tfrac12\delta(P_N,Q_N)^2] > \frac{1}{N}\log(\frac{1+\varepsilon}{2})$$

voor voldoend grote N, zodat

$$-\tfrac{1}{2}\delta(P_N,Q_N)^2 + O(\delta(P_N,Q_N)^4) > \log\frac{1+\varepsilon}{2}.$$

Hieruit volgt dat $\delta(P_N,Q_N) = O(N^{-\frac{1}{2}})$ als $N \to \infty$. Toepassing van lemma 1.11 geeft dat $\| P_N - Q_N \| = O(N^{-\frac{1}{2}})$ voor $N \to \infty$, waarmee a) bewezen is.

b) In punt a) hebben we de implicatie $\{P_N^N\} \not\sim \{Q_N^N\} \Rightarrow \| P_{N_1} - Q_N \| = O(N^{-\frac{1}{2}})$ voor $N \to \infty$ bewezen. Laat $\lim_{N\to\infty} \| P_N - Q_N \| = 0$ én laat $N^{\frac{1}{2}} \| P_N - Q_N \| \to \infty$. Dan kan $\{P_N^N\}$ en $\{Q_N^N\}$ dus niet naburig zijn. Als voorbeelden noemen we:

(i) voor $N = 1,2,\ldots$ is P_N de homogene verdeling op $(0,1)$ en Q_N de homogene verdeling op $(a_N, 1+a_N)$ met $\lim_{N\to\infty} a_N = 0$ en $\lim_{N\to\infty} N^{\frac{1}{2}} a_N = \infty$,

(ii) voor $N = 1,2,\ldots$, $P_N = N(\mu_N;\sigma_N^2)$ en $Q_N = N(\nu_N;\sigma_N^2)$ met $\lim_{N\to\infty} \frac{\mu_N - \nu_N}{\sigma_N} = 0$ en $\lim N(\frac{\mu_N-\nu_N}{\sigma_N})^2 = \infty$ (zie opgave 1.5).

Opmerking: Er is ook een geheel ander bewijs van punt a) in opgave 1.7 mogelijk, dat berust op stelling 1.13. Uit de naburigheid van $\{P_N^N\}$ en $\{Q_N^N\}$ volgt dat

$$L(\Lambda_N) = L\Big(\sum_{i=1}^{N} \log\frac{p_N(X_i)}{q_N(X_i)}\Big)$$

relatief compact is, zowel onder $\{P_N\}$ als onder $\{Q_N\}$. Maar dan (zie Gnedenko en Kolmogorov [9], pag. 58):

$$\log\frac{p_N(X)}{q_N(X)} \to 0 \qquad \text{zowel in } P_N\text{-kans als in } Q_N\text{-kans en dus}$$

ook in $\mu_N = \frac{P_N+Q_N}{2}$ - kans, zodat

$$(*) \quad \frac{p_N(X)}{q_N(X)} \to_{\mu_N} 1, \quad \text{d.w.z.} \quad \frac{p_N(X)-q_N(X)}{q_N(X)} \to_{\mu_N} 0.$$

Zij $\varepsilon > 0$ willekeurig, dan is

$$\|P_N-Q_N\| = \tfrac{1}{2}\int |p_N-q_N|\,d\mu_N = \tfrac{1}{2}\int \chi(\left|\frac{p_N-q_N}{q_N}\right| \geq \varepsilon)|p_N-q_N|\,d\mu_N +$$

$$+ \tfrac{1}{2}\int \chi(\left|\frac{p_N-q_N}{q_N}\right| \leq \varepsilon)|p_N-q_N|\,d\mu_N \leq \tfrac{1}{2}\int \chi(\left|\frac{p_N-q_N}{q_N}\right| \geq \varepsilon)(p_N+q_N)\,d\mu_N +$$

$$+ \tfrac{1}{2}\int \varepsilon q_N\,d\mu_N = \int \chi(\left|\frac{p_N-q_N}{q_N}\right| \geq \varepsilon)\,d\mu_N + \tfrac{1}{2}\varepsilon.$$

Wegens (*) hebben we dus $0 \leq \lim\sup_{N\to\infty} \|P_N-Q_N\| \leq \frac{1}{2}\varepsilon$ voor elke $\varepsilon > 0$ en derhalve $\|P_N-Q_N\| \to 0$ voor $N \to \infty$.

1.8. a) Zij $\{P^N\} \not\Leftrightarrow \{Q^N\}$. Laat voor $N = 1,2,\ldots$ $A_N \in \mathcal{A}_N$ zodat $\lim_{N\to\infty} P_N(A_N) = 0$. Dan geldt tevens dat $\lim_{N\to\infty} P^N(X_1\times X_2\times\ldots\times X_{N-1}\times A_N) = 0$. Hieruit volgt dat

$$\lim_{N\to\infty} Q_N(A_N) = \lim_{N\to\infty} Q^N(X_1\times\ldots\times X_{N-1}\times A_N) = 0.$$

Evenzo impliceert $\lim_{N\to\infty} Q_N(A_N) = 0$ dat $\lim_{N\to\infty} P_N(A_N) = 0$, zodat $\{P_N\} \not\Leftrightarrow \{Q_N\}$.
Stel dat P_N niet equivalent is met Q_N voor alle N. Dan is er een index n en een verzameling $A_n \in \mathcal{A}_n$ zodat $P_n(A_n) = 0$ en $Q_n(A_n) > 0$ (of omgekeerd). Maar hieruit volgt dat voor alle $N = n,n+1,\ldots,$

$$P^N(X_1\times\ldots\times X_{n-1}\times A_n\times X_{n+1}\times\ldots\times X_N) = 0$$

en $Q^N(X_1\times\ldots\times X_{n-1}\times A_n\times X_{n+1}\times\ldots\times X_N) = Q_n(A_n) = 0$, hetgeen in tegenspraak is met het gegeven.

Opmerking: Men kan zelfs op dezelfde wijze als in de oplossing van opgave 1.7 laten zien dat uit het gegeven volgt dat $\lim_{N\to\infty} \|P_N-Q_N\| = 0$.

b) Beschouw het volgende tegenvoorbeeld.
Voor $N = 1,2,\ldots$ is P_N de homogene verdeling op (0,1) en heeft Q_N t.o.v. de Lebesgue maat de dichtheid

$$q_N(x) = \begin{cases} \frac{1}{2}N & \text{voor } 0 < x < 1/N \\ \\ \dfrac{N}{2(N-1)} & \text{voor } 1/N < x < 1. \end{cases}$$

Kies voor $N = 1,2,\ldots, A_N = (0,1/N) \times \Pi_{i=2}^{N} (0,1)$. Dan is $\lim_{N\to\infty} P^N(A_N) = \lim_{N\to\infty} 1/N = 0$ en $\lim_{N\to\infty} Q^N(A_N) = \lim_{N\to\infty} \frac{1}{2} = \frac{1}{2}$, zodat $\{P^N\}$ en $\{Q^N\}$ niet naburig zijn.

1.9. Kies voor P_N de homogene verdeling op $(0,1)$, $N = 1,2,\ldots$ en kies voor Q_N de homogene verdeling op $(0,\frac{1}{2})$, $N = 1,2,\ldots$.
Dan is $Q_N(L_N=2) = 1$ voor alle N, zodat triviaal aan (b') voldaan is. Voorts is $P_N(L_N=2) = P_N(L_N=0) = \frac{1}{2}$ voor alle N, zodat ook aan (c') is voldaan. Daar $P_N(\Lambda_N=-\infty) = P_N(L_N=0) = \frac{1}{2}$ voor alle N, is $\{\mathcal{L}(\Lambda_N|P_N)\}$ niet uniform beperkt en dus niet relatief compact, zodat op grond van stelling 1.13 niet geldt $\{P_N\} \lhd\rhd \{Q_N\}$.
Uiteraard kan men legio rijen kansmaten $\{P_N\}$ en $\{Q_N\}$ construeren zódat $P_N(\Lambda_N = -\infty) = P_N(L_N=0)$ niet naar 0 convergeert, terwijl wel aan (b') en (c') is voldaan.

1.10. a) $\Rightarrow$ b). Zij $\{P_N\} \lhd\rhd \{Q_N\}$ en zij $\{\mathcal{L}(T_N|P_N)\}$ relatief compact. Dan is $\{\mathcal{L}(T_N|P_N)\}$ ook uniform beperkt. Zij $\{\varepsilon_i\}$ een rij reële getallen zó dat $\varepsilon_i \downarrow 0$. Dan bestaat er een rij $\{c_i\}$ zó dat $c_i \uparrow \infty$ én $P_N(|T_N| > c_i) \le \varepsilon_i$ voor alle $N = 1,2,\ldots$ en $i = 1,2,\ldots$.
Onderstel dat $\{\mathcal{L}(T_N|Q_N)\}$ niet relatief compact en dus niet uniform beperkt is. Dan bestaat er een $\varepsilon > 0$ en een deelrij $\{N_i\}$ zó dat $Q_{N_i}(|T_{N_i}| > c_i) \ge \varepsilon$ voor $i = 1,2,\ldots$. Zij $A_{N_i} := [|T_{N_i}| > c_i]$ voor $i = 1,2,\ldots$, dan volgt voor $i \to \infty$ dat $P_{N_i}(A_{N_i}) \to 0$ én $Q_{N_i}(A_{N_i}) \not\to 0$. Dan zijn $\{P_{N_i}\}$ en $\{Q_{N_i}\}$ dus niet naburig, hetgeen impliceert dat $\{P_N\}$ en $\{Q_N\}$ evenmin naburig zijn, in strijd met het gegeven. Maar dan is $\{\mathcal{L}(T_N|Q_N)\}$ dus wel relatief compact. Analoog bewijst men de implicatie $\{\mathcal{L}(T_N|Q_N)\}$ relatief compact $\Rightarrow \{\mathcal{L}(T_N|P_N)\}$ relatief compact.

b) $\Rightarrow$ a). Laat de equivalentie ($\{\mathcal{L}(T_N|P_N)\}$ is relatief compact) $\Longleftrightarrow$

($\{L(T_N|Q_N)\}$ is relatief compact) gelden voor alle A_N-meetbare stochastische grootheden T_N(N=1,2,...). In het bijzonder geldt deze equivalentie dus voor het aannemelijksquotient L_N(N=1,2,...). Daar $\{L(L_N|P_N)\}$ altijd relatief compact is (zie pag. 19) volgt dat ook $\{L(L_N|Q_N)\}$ relatief compact is en dus m.b.v. stelling 1.14 dat $\{Q_N\} \lhd \{P_N\}$. Door verwisseling van de rol van P_N en Q_N bewijst men analoog dat $\{P_N\} \lhd \{Q_N\}$.

1.11. Voor het gemak voeren we de volgende notaties in:

$$\tilde{\mu} = (\mu_1,\ldots,\mu_k);\ \tilde{\zeta} = (\sigma_{01},\ldots,\sigma_{0k});\ \zeta = (\sigma^2,\tilde{\zeta}) \text{ en } \tilde{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{12} & & \ddots & \\ \vdots & & & \\ \sigma_{1k} & & & \sigma_{kk} \end{pmatrix}.$$

Uit het gegeven volgt m.b.v. lemma 1.7 dat $L(\Lambda_N|P_N) \to_z N(-\frac{1}{2}\sigma^2;\sigma^2)$, zodat $\{P_N\} \Leftrightarrow \{Q_N\}$ (lemma 1.17). Toepassing van stelling 1.15 levert dat $L((\Lambda_N,T_N)|Q_N) \to_z L'$,
waarbij

$$\frac{dL'}{dN(\mu,\Sigma)}(y,t) = \exp(y) \qquad (y\in\mathbb{R}, t\in\mathbb{R}^k).$$

Zij $\sigma^2 = 0$ (en dus $\Sigma_1 = 0$), dan is $\exp(y) = 1$ b.o. onder $N(\mu,\Sigma)$, zodat in dit geval onmiddellijk volgt dat $L((\Lambda_N,T_N)|Q_N) \to_z N(\mu;\Sigma)$. Veronderstel dus vervolgens dat $\sigma^2 > 0$. Zij $\tilde{\phi}$ de karakteristieke functie van L', dan is dus

$$\tilde{\phi}(s) = \int \exp(is_0y+i\tilde{s}t'+y)dN(\mu;\Sigma)(y,t),$$

waarin $s = (s_0,\tilde{s})$ en $\tilde{s} = (s_1,\ldots,s_k)$. We berekenen deze integraal door eerst de integrand te integreren naar t met de voorwaardelijke verdeling gegeven y en vervolgens te integreren naar y met de marginale verdeling van y. Zij L'_y de (eventueel ontaarde) voorwaardelijke verdeling gegeven y, dan is L'_y de normale $N(\tilde{\mu}_y;\tilde{\tilde{\Sigma}})$ verdeling, waarin $\tilde{\mu}_y = \tilde{\mu} + \tilde{\zeta}\sigma^{-2}(y+\sigma^2/2)$ en $\tilde{\tilde{\Sigma}} = \tilde{\Sigma} - \tilde{\zeta}'\sigma^{-2}\tilde{\zeta}$ (zie ANDERSON [1]).

De marginale verdeling van y is de normale $N(-\frac{1}{2}\sigma^2;\sigma^2)$ verdeling zodat geldt

$$\tilde{\phi}(s) = \int_{\mathbb{R}} (2\pi\sigma^2)^{-\frac{1}{2}}\exp(is_0y+y-\tfrac{1}{2}\sigma^{-2}(y+\tfrac{1}{2}\sigma^2)^2\Big[\int_{\mathbb{R}^k} \exp(i\tilde{s}t')d\tilde{L}'_y(t)\Big]dy.$$

Daar de vorm tussen rechte haken juist de karakteristieke functie is van de L'_y-verdeling, geldt

$$\int_{\mathbb{R}^k} \exp(i\tilde{s}t')d\tilde{L}'_y(t) = \exp(i\tilde{\tilde{s}}\mu'_y-\tfrac{1}{2}\tilde{\tilde{s}}\tilde{\Sigma}\tilde{\tilde{s}}') =$$

$$= \exp(i\tilde{\tilde{s}}\mu'+\tfrac{1}{2}i\tilde{\tilde{s}}\zeta'+i\sigma^{-2}\tilde{\tilde{s}}\zeta'y-\tfrac{1}{2}\tilde{\tilde{s}}\tilde{\Sigma}\tilde{\tilde{s}}'),$$

zodat

$$\tilde{\phi}(s) = \exp(i\tilde{\tilde{s}}\mu'+\tfrac{1}{2}i\tilde{\tilde{s}}\zeta'-\tfrac{1}{2}\tilde{\tilde{s}}\tilde{\Sigma}\tilde{\tilde{s}}').$$

$$\cdot \int_{\mathbb{R}} (2\pi\sigma^2)^{-\frac{1}{2}}\exp(y+is_0y+i\sigma^{-2}\tilde{\tilde{s}}\zeta'y-\tfrac{1}{2}\sigma^{-2}(y+\tfrac{1}{2}\sigma^2)^2)dy.$$

We kunnen eenvoudig narekenen dat de integraal in deze laatste uitdrukking gelijk is aan

$$\int_{\mathbb{R}} (2\pi\sigma^2)^{-\frac{1}{2}}\exp(-\tfrac{1}{2}\sigma^{-2}(y-(i\tilde{\tilde{s}}\zeta'+\sigma^2(\tfrac{1}{2}+is_0)))^2)dy.$$

$$\cdot \exp\{(-\frac{1}{8}\sigma^2)+\tfrac{1}{2}\sigma^{-2}(i\tilde{\tilde{s}}\zeta'+\sigma^2(\tfrac{1}{2}+is_0))^2\} =$$

$$= \exp(-\frac{1}{8}\sigma^2+\tfrac{1}{2}\sigma^{-2}(i\tilde{\tilde{s}}\zeta'+\sigma^2(\tfrac{1}{2}+is_0))^2),$$

zodat

$$\tilde{\phi}(s) = \exp(i\tilde{\tilde{s}}\mu'+\tfrac{1}{2}i\tilde{\tilde{s}}\zeta'-\tfrac{1}{2}\tilde{\tilde{s}}\tilde{\Sigma}\tilde{\tilde{s}}' - \frac{1}{8}\sigma^2-\tfrac{1}{2}\sigma^{-2}(\tilde{\tilde{s}}\zeta')^2 +$$

$$+ \tfrac{1}{2}i\tilde{\tilde{s}}\zeta' - s_0\tilde{\tilde{s}}\zeta' + \frac{1}{8}\sigma^2 + \tfrac{1}{2}is_0\sigma^2 - \tfrac{1}{2}\sigma^2s_0),$$

hetgeen juist de karakteristieke functie is van de normale $N(\nu;\Sigma)$ verdeling.

1.12. We zullen laten zien dat voor elke vector $c \in \mathbb{R}^k$ geldt dat

$$L((cT_N'-\Lambda_N)|P_N) \to_z N(\tfrac{1}{2}\sigma^2;(-1,c)\begin{pmatrix}\sigma^2 & h\Gamma\\ \Gamma h' & \Gamma\end{pmatrix}(-1,c)').$$

Het verlangde resultaat volgt dan direct m.b.v. lemma 1.7. Nu is

$cT_N' - \Lambda_N = hT_N' - \Lambda_N + (c-h)T_N'$, waarin

$$hT_N' - \Lambda_N \to_{P_N} \tfrac{1}{2}\sigma^2 \text{ en } L((c-h)T_N'|P_N) \to_z N(0;(c-h)\Gamma(c-h)'),$$

zodat inderdaad

$$L((cT_N'-\Lambda_N)|P_N) \to_z N(\tfrac{1}{2}\sigma^2;(c-h)\Gamma(c-h)' = N(\tfrac{1}{2}\sigma^2;(-1,c)\cdot$$
$$\cdot\begin{pmatrix}\sigma^2 & h\Gamma\\ \Gamma h' & \Gamma\end{pmatrix}(-1,c)').$$

Vraagstukken

2.1. Laten voor $N = 1,2,\ldots$, de onafhankelijke componenten X_{Nn} van de stochastische vector X_N een normale $N(\zeta(N,n);1)$ verdeling bezitten $(n=1,2,\ldots,N)$.
Beschouw het toetsingsprobleem $(\alpha;\theta(N),\tau(N))$, waarin $0 < \alpha < 1$, $\theta(N,n) = \theta_0$ voor $n = 1,2,\ldots,N$, $\tau(N,n) = \tau_{1N}$ voor $n = 1,2,\ldots,N_1$, $\tau(N,n) = \tau_{2N}$ voor $n = N_1+1,\ldots,N$. Hierin is $N = N_1+N_2$, θ_0 een vaste constante en $\tau_{1N} \neq \tau_{2N}$. Zie voorbeeld 2.3.
Ga na dat de limiet van het onderscheidingsvermogen in $\tau(N)$ van de meest onderscheidende toets strikt tussen α en 1 ligt bij de keuze

$$\tau_{1N} = \theta_0 - \eta N_1^{-\frac{1}{2}} N_2^{\frac{1}{2}} N^{-\frac{1}{2}} + o(N_1^{-\frac{1}{2}}),$$
$$\tau_{2N} = \theta_0 + \eta N_1^{\frac{1}{2}} N_2^{-\frac{1}{2}} N^{-\frac{1}{2}} + o(N_2^{-\frac{1}{2}}),$$

voor willekeurige, doch vaste $\eta \neq 0$, als $N_1, N_2 \to \infty$.

2.2. Laten voor $N = 1,2,\ldots$, de onafhankelijke componenten X_{Nn} van de stochastische vector X_N een 2-dim. normale $N(0,0;1,1,\zeta(N,n))$ verdeling bezitten met $-1 < \zeta(N,n) < 1$, $(n=1,2,\ldots,N)$. Zie voorbeeld 2.2.
Ga na dat de limiet van het onderscheidingsvermogen in $\tau(N)$ van de meest onderscheidende toets voor de nulhypothese $\theta(N,n) = 0$ (alle n,N) tegen $\tau(N)$ strikt tussen α en 1 ligt bij de keuze

$$\tau(N,n) = \eta N^{-\frac{1}{2}} + o(N^{-\frac{1}{2}}), \qquad n = 1,2,\ldots,N,$$

voor vaste doch willekeurige $\eta \neq 0$, als $N \to \infty$.

2.3. Laten $X_1, X_2, \ldots$, onafhankelijke $N(a;\sigma_0^2)$ verdeelde stochastische variabelen zijn met $a \in \mathbb{R}$ en bekende $\sigma_0^2 > 0$. Laat $\{P_a, a \in \mathbb{R}\}$ de bijbehorende familie van kansmaten zijn op $(\mathbb{R},\mathcal{B})$ en laat voor iedere $a \in \mathbb{R}$, g_a de dichtheid van P_a t.o.v. de Lebesgue-maat zijn op $(\mathbb{R},\mathcal{B})$.

a) Ga na dat deze familie van kansmaten voldoet aan voorwaarde (I) (hoofdstuk 2, pag. 45) en bereken de Fischer-informatie in een vast punt $a_0 \in \mathbb{R}$ van de familie kansdichtheden $\{g_a, a\in\mathbb{R}\}$.

b) Beschouw voor N = 1,2,..., vervolgens de productmaat $P_{\theta(N)}(P_{\tau(N)})$ van de stochastische vector $X_N = (X_{N1},\ldots,X_{NN})$, waarvan de componenten X_{Nn} onafhankelijk zijn en identiek verdeeld volgens $P_{\theta(N,n)}(P_{\tau(N,n)})$ voor n = 1,2,...,N.
Kies

$$\theta(N,n) = a_0 \quad \text{voor } n = 1,2,\ldots,N, \qquad N = 1,2,\ldots,$$

en

$$\tau(N,n) = a_0 + \eta N^{-\frac{1}{2}} \quad \text{voor } n = 1,2,\ldots,N, \quad N = 1,2,\ldots,$$

en zekere η, $0 < \eta < \infty$.

Bewijs m.b.v. stelling 2.1 dat de Gauss-toets asymptotisch meest onderscheidend is. Laat vervolgens zien dat aan de voorwaarden van stelling 2.5 (pag. 45) is voldaan en ga na dat de AMP toets die men aldus verkrijgt gelijk is aan de Gauss-toets.

2.4. Zij $\{P_\zeta, \zeta\in\mathbb{R}\}$ de familie dubbel-exponentiële verdelingen op $(\mathbb{R},\mathcal{B})$ met dichtheden

$$f_\zeta(x) = \frac{1}{2}\exp(-|x-\zeta|),$$

t.o.v. de Lebesgue-maat op $\mathbb{R}$.
Ga na dat deze familie dichtheden aan voorwaarde (I) (hoofdstuk 2, pag. 45) voldoet bij $\theta_0 = 0$.
Zij voor N = 1,2,..., $X_{N1},\ldots,X_{NN}$ een aselecte steekproef uit bovenstaande dubbel-exponentiële verdeling.
Konstrueer een AMP toets voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ met $\theta(N) = 0$ (alle N) en $\tau(N,n) = \tau_N > 0$, n = 1,2,...,N, N = 1,2,..., met $\tau_N = \eta N^{-\frac{1}{2}}$ voor zekere η, $0 < \eta < \infty$.

2.5. Zij $\{P_\zeta, \zeta \in \mathbb{R}\}$ de familie Cauchy-verdelingen op $(\mathbb{R}, \mathcal{B})$ met dichtheden

$$f_\zeta(x) = \frac{1}{\pi} \frac{1}{1+(x-\zeta)^2},$$

t.o.v. de Lebesgue-maat op $\mathbb{R}$.
Ga na dat deze familie dichtheden aan voorwaarde (I) (hoofdstuk 2, pag. 45) voldoet bij $\theta_0 = 0$ en bepaal m.b.v. stelling 2.5 (pag. 45) een AMP toets.

2.6. Zij $\{P_\zeta, 0 \leq \zeta < 1\}$ de familie van homogene kansverdelingen op $(\zeta, 1)$ met dichtheden

$$f_\zeta(x) = (1-\zeta)^{-1} \chi_{(\zeta,1)}(x),$$

t.o.v. de Lebesgue-maat op $\mathbb{R}$. Zij $\theta(N,n) = \theta_0 = 0$ en $0 < \tau(N,n) = \tau_N < 1$ voor $n = 1,2,\ldots,N$, $N = 1,2,\ldots$.
a) Bewijs de equivalentie:

$$\left.\begin{array}{l} \lim\limits_{N\to\infty} \max\limits_{1\leq n\leq N} [\delta(P_{\theta(N,n)}, P_{\tau(N,n)})]^2 = 0, \\ \\ \lim\limits_{N\to\infty} \sum\limits_{n=1}^{N} [\delta(P_{\theta(N,n)}, P_{\tau(N,n)})]^2 = a^2, \end{array}\right\} \Leftrightarrow$$

$$\Leftrightarrow \tau_N = N^{-1}a^2 + o(N^{-1}), \quad N\to\infty ,$$

als hierin a voldoet aan $0 < a < \infty$.

b) Laat zien dat voorwaarde (I) (hoofdstuk 2, pag. 45) niet vervuld is.

2.7. Beschouw het asymptotische toetsingsprobleem $(\alpha; \{\theta(N)\}, \{\tau(N)\})$ voor $0 < \alpha < 1$ en laat $\{\phi_N\}$ een asymptotisch meest onderscheidende rij toetsen zijn voor dit probleem.

Bewijs de volgende implicaties:

a) $\lim_{N\to\infty} \int \phi_N \, dP_{\tau(N)} = \alpha \Rightarrow \{P_{\theta(N)}\} \triangleleft\triangleright \{P_{\tau(N)}\}$,

b) $\lim_{N\to\infty} \int \phi_N \, dP_{\tau(N)} = 1 \Rightarrow \{P_{\theta(N)}\} \not\triangleleft \{P_{\tau(N)}\}$.

2.8. Laat voor $N = 1,2,\ldots$, $X_{N1},X_{N2},\ldots,X_{NN}$ onafhankelijk normaal $N(\zeta;1)$ verdeeld zijn.
Bepaal de ARE van de tekentoets t.o.v. de Gauss-toets voor het asymptotische toetsingsprobleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ met $0 < \alpha < 1$, $\theta(N,n) = 0$ voor alle n,N en $\tau(N,n) = \eta N^{-\frac{1}{2}}$ voor $n = 1,2,\ldots,N$, $N = 1,2,\ldots$, voor zekere η, $0 < \eta < \infty$.

2.9. Laat voor $N = 1,2,\ldots$, $X_{N1},X_{N2},\ldots,X_{NN}$ onafhankelijk normaal verdeeld zijn met dezelfde variantie $\sigma^2 (0<\sigma^2<\infty)$.
Zij voor alle n,N, $\zeta(N,n) = EX_{Nn}$.
Bepaal voor vaste σ^2 de ARE van de twee-steekproeven t-toets t.o.v. de twee-steekproeven Gauss-toets voor het asymptotische toetsingsprobleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$, waarin $\theta(N,n) = \theta_0$ voor alle $n,N(\theta_0 \in \mathbb{R}$, onbekend), $\tau(N,n) = \theta_0 + \eta N^{\frac{1}{2}} N_1^{-\frac{1}{2}} N_2^{-\frac{1}{2}}$ voor $n = 1,2,\ldots,N_1$, en $\tau(N,n) = \theta_0$ voor $n = N_1+1,\ldots,N$ voor zekere η, $0 < \eta < \infty$, $N = 1,2,\ldots$, als zowel $N_1 \to \infty$ als $N_2 \to \infty$.

Oplossingen der vraagstukken

2.1. Om de toetsingsgrootheid van de meest onderscheidende toets te vinden passen we het lemma van Neyman-Pearson toe:

$$\frac{p_{\tau(N)}(X_N)}{p_{\theta(N)}(X_N)} = \frac{\exp\{-\frac{1}{2}\sum_{n=1}^{N_1}(X_{Nn}-\tau_{1N})^2 - \frac{1}{2}\sum_{n=N_1+1}^{N}(X_{Nn}-\tau_{2N})^2\}}{\exp\{-\frac{1}{2}\sum_{n=1}^{N}(X_{Nn}-\theta_0)^2\}} =$$

$$= \exp\{(\tau_{1N}-\theta_0)\sum_{n=1}^{N_1}X_{Nn} + (\tau_{2N}-\theta_0)\sum_{n=N_1+1}^{N}X_{Nn} + \text{constante}\}.$$

Deze meest onderscheidende toets verwerpt H_0 voor grote waarden van $T_N := N_1(\tau_{1N}-\theta_0)\bar{X}^{(1)} + N_2(\tau_{2N}-\theta_0)\bar{X}^{(2)}$, waarbij $\bar{X}^{(1)} = N_1^{-1}\sum_{n=1}^{N_1}X_{Nn}$ en $\bar{X}^{(2)} = N_2^{-1}\sum_{n=N_1+1}^{N}X_{Nn}$.
Onder H_0 heeft T_N een normale $N(N_1(\tau_{1N}-\theta_0)\theta_0+N_2(\tau_{2N}-\theta_0)\theta_0; N_1(\tau_{1N}-\theta_0)^2 + N_2(\tau_{2N}-\theta_0)^2)$ verdeling, zodat de kritieke waarde wordt:

$$N_1(\tau_{1N}-\theta_0)\theta_0 + N_2(\tau_{2N}-\theta_0)\theta_0 + \xi_{1-\alpha}\left[N_1(\tau_{1N}-\theta_0)^2+N_2(\tau_{2N}-\theta_0)^2\right]^{\frac{1}{2}}.$$

Onder H_1 heeft T_N een normale $N(N_1(\tau_{1N}-\theta_0)\tau_{1N}+N_2(\tau_{2N}-\theta_0)\tau_{2N}; N_1(\tau_{1N}-\theta_0)^2 + N_2(\tau_{2N}-\theta_0)^2)$ verdeling, zodat het onderscheidingsvermogen in τ_N gelijk is aan

$$\beta(\tau_N) = 1 - \Phi(\xi_{1-\alpha} - \left[N_1(\tau_{1N}-\theta_0)^2+N_2(\tau_{2N}-\theta_0)^2\right]^{\frac{1}{2}}).$$

In deze uitdrukking substitueren we nu

$$\tau_{1N} = \theta_0 - \eta N_1^{-\frac{1}{2}}N_2^{\frac{1}{2}}N^{-\frac{1}{2}} + o(N_1^{-\frac{1}{2}}),$$
$$\tau_{2N} = \theta_0 + \eta N_1^{\frac{1}{2}}N_2^{-\frac{1}{2}}N^{-\frac{1}{2}} + o(N_2^{-\frac{1}{2}}),$$

zodat

$$\beta(\tau_N) = 1 - \Phi(\xi_{1-\alpha} - (\eta^2+o(1))^{\frac{1}{2}}), \qquad \text{voor } N_1, N_2 \to \infty,$$

en dus

$$\beta(\tau_N) \to 1 - \Phi(\xi_{1-\alpha} - |\eta|), \qquad \text{voor } N_1, N_2 \to \infty.$$

2.2. Wij passen wederom het lemma van Neyman-Pearson toe.
Zij $X_{Nn} = (Y_{Nn}, Z_{Nn})$, $n = 1,\ldots,N$; $N = 1,2,\ldots$, zij $\theta(N) = (0,\ldots,0)$ en $\tau(N) = (\tau_N,\ldots,\tau_N)$, dan is

$$\frac{p_{\tau(N)}(X_N)}{p_{\theta(N)}(X_N)} = \frac{\exp\{-\frac{1}{2}(1-\tau_N^2)^{-1} \sum_{n=1}^{N} [Y_{Nn}^2 + Z_{Nn}^2 - 2\tau_N Y_{Nn} Z_{Nn}]\}}{\exp\{-\frac{1}{2} \sum_{n=1}^{N} [Y_{Nn}^2 + Z_{Nn}^2]\}} =$$

$$= \exp\{-\tfrac{1}{2}\tau_N^2(1-\tau_N^2)^{-1} \sum_{n=1}^{N} [Y_{Nn}^2 + Z_{Nn}^2 - 2\tau_N^{-1} Y_{Nn} Z_{Nn}]\}.$$

Bij de meest onderscheidende toets van H_0 : $\zeta(N,n) = (0,\ldots,0)$, $n = 1,2,\ldots,N$, tegen $\zeta(N,n) = (\tau_N,\ldots,\tau_N)$, $n = 1,2,\ldots,N$, wordt H_0 dus verworpen als

$$T_N := \sum_{n=1}^{N} (Y_{Nn}^2 + Z_{Nn}^2 - 2\tau_N^{-1} Y_{Nn} Z_{Nn}) < c_{N,\alpha}.$$

We bepalen eerst een asymptotische uitdrukking voor $c_{N,\alpha}$. Onder H_0 is

$$E_{\theta(N)}\, T_N = 2N,$$
$$Var_{\theta(N)}\, T_N = N\, E_0(Y_{Nn}^2 + Z_{Nn}^2 - 2\tau_N^{-1} Y_{Nn} Z_{Nn})^2 - 4N = 4N(1+\tau_N^{-2}).$$

Uit de centrale limietstelling leiden we nu eenvoudig af dat

$$L((T_N - 2N)\tfrac{1}{2}N^{-\frac{1}{2}}(1+\tau_N^{-2})^{-\frac{1}{2}} \mid P_{\theta(N)}) \to_z N(0;1),$$

mits $\lim\sup_{N\to\infty} |\tau_N| < 1$, zodat onder deze voorwaarde

$$c_{N,\alpha} = 2N + 2N^{\frac{1}{2}}(1+\tau_N^{-2})^{\frac{1}{2}}\xi_\alpha + o(N^{\frac{1}{2}}), \qquad \text{als } N \to \infty.$$

Voorts geldt onder de alternatieve hypothese

$$E_{\tau(N)}\ T_N = N\ E_{\tau_N}(Y_{Nn}^2+Z_{Nn}^2-2\tau_N^{-1}Y_{Nn}Z_{Nn}) = 0$$

en

$$\begin{aligned} Var_{\tau(N)}\ T_N &= N\ E_{\tau_N}(Y_{Nn}^2+Z_{Nn}^2-2\tau_N^{-1}Y_{Nn}Z_{Nn})^2 = \\ &= N\{2E_{\tau_N}Y_{Nn}^4+2(1+2\tau_N^{-2})E_{\tau_N}Y_{Nn}^2Z_{Nn}^2-8\tau_N^{-1}E_{\tau_N}Y_{Nn}Z_{Nn}^3\} = \\ &= N\{6+2(1+2\tau_N^{-2})(1+2\tau_N^2)-8\tau_N^{-1}(3\tau_N)\} = \\ &= 4N\tau_N^{-2}(1-\tau_N^2)^2 \end{aligned}$$

(de gemengde momenten bepaalt men wellicht het eenvoudigst door herhaalde partiële differentiatie van de karakteristieke funktie $\exp\{-\frac{1}{2}(t_1^2+2\tau_N t_1 t_2+t_2^2)\}$ van (Y_{Nn},Z_{Nn})).

Ook thans kan men met de centrale limietstelling laten zien dat

$$L(T_N\tfrac{1}{2}N^{-\frac{1}{2}}|\tau_N|(1-\tau_N^2)^{-1} \mid P_{\tau(N)}) \to_z N(0;1),$$

mits $\lim\sup_{N\to\infty}|\tau_N| < 1$. Hieruit volgt dat onder deze voorwaarde het onderscheidingsvermogen $\beta(\tau_N)$ van de toets voor $N \to \infty$ gelijk is aan

$$\begin{aligned} P_{\tau(N)}(T_N \le c_{N,\alpha}) &= \Phi(c_{N,\alpha}\ \tfrac{1}{2}N^{-\frac{1}{2}}|\tau_N|(1-\tau_N^2)^{-1}) + o(1) = \\ &= \Phi(\xi_\alpha(1+\tau_N^2)^{\frac{1}{2}}(1-\tau_N^2)^{-1}+N^{\frac{1}{2}}|\tau_N|(1-\tau_N^2)^{-1}+o(1))+o(1), \end{aligned}$$

zodat, als $\tau_N = \eta N^{-\frac{1}{2}} + o(N^{-\frac{1}{2}})$,

$$\lim_{N\to\infty}\beta(\tau_N) = \Phi(\xi_\alpha+|\eta|).$$

2.3. **a)** We kiezen voor μ de Lebesgue-maat op $(\mathbb{R},\mathcal{B})$ en $X_0 = \mathbb{R}$. Nu geldt $g_0(x) = (2\pi)^{-\frac{1}{2}}\sigma_0^{-1}\ \exp(-\frac{1}{2}\sigma_0^{-2}(x-a_0)^2) > 0$ op $\mathbb{R}$, en

$$\frac{\partial}{\partial a}\ g_a(x) = g_a(x)(x-a)\sigma_0^{-2},$$

zodat $\dot{g}_0(x) = g_0(x)(x-a_0)\sigma_0^{-2}$ bestaat voor alle $x \in X_0$.
Vervolgens verifiëren we (2.18):

$$\int \dot{g}_0(x)dx = \sigma_0^{-2} \int g_0(x)(x-a_0)dx = \sigma_0^{-2}(E_0(X)-a_0) = 0,$$

en

$$\int (\dot{g}_0^2(x)/g_0(x))dx = \int (g_0^2(x)(x-a_0)^2\sigma_0^{-4}/g_0(x))dx =$$

$$= \sigma_0^{-4} \int g_0(x)(x-a_0)^2dx = \sigma_0^2\,\sigma_0^{-4} = \sigma_0^{-2} = J_0,$$

zodat $0 < J_0 < \infty$.
Aan (2.19) is ook voldaan daar

$$\lim_{a\to a_0} \int \left((g_a(x))^{\frac{1}{2}} - (g_0(x))^{\frac{1}{2}}\right)^2 (a-a_0)^{-2}\,dx =$$

$$= \lim_{a\to a_0} (a-a_0)^{-2}\Big(\int (g_a(x)+g_0(x))dx - 2(2\pi)^{-\frac{1}{2}}\sigma_0^{-1}$$

$$\cdot \int \exp\Big(-(4\sigma_0^2)^{-1}\Big[(x-a)^2+(x-a_0)^2\Big]\Big)dx\Big) =$$

$$= \lim_{a\to a_0} 2(a-a_0)^{-2}\Big[1-(2\pi)^{-\frac{1}{2}}\sigma_0^{-1} \int \exp\Big(-(2\sigma_0^2)^{-1}(x-(a+a_0)/2)^2 -$$

$$- (2\sigma_0^2)^{-1}[(a^2+a_0^2)/2-((a+a_0)/2)^2]\Big)dx\Big] =$$

$$= 2 \lim_{a\to a_0} (a-a_0)^{-2}\Big(1-\exp(-(8\sigma_0^2)^{-1}(a-a_0)^2)\Big) =$$

$$= 2 \lim_{a\to a_0} (a-a_0)^{-2}((8\sigma_0^2)^{-1}(a-a_0)^2+O(a-a_0)^4) = (4\sigma_0^2)^{-1} = J_0/4.$$

b) Daar de Gauss-toets de meest onderscheidende toets is voor het probleem $(\alpha;\theta(N),\tau(N))$, voor $N = 1,2,\ldots$, is de Gauss-toets volgens stelling 2.1 ook asymptotisch meest onderscheidend.
Aan de voorwaarden van stelling 2.5 is voldaan daar deze familie van kansmaten voldoet aan voorwaarde (I) en bovendien geldt dat $\max_{1\le n\le N}[\tau(N,n)-a_0]^2 = \eta^2N^{-1} \to 0$ en $\sum_{n=1}^{N}[\tau(N,n)-a_0]^2 = \sum_{n=1}^{N} \eta^2N^{-1} = \eta^2 \to \eta^2$, voor $N \to \infty$.

In deze situatie is S_{Nn} (zie stelling 2.5) gelijk aan

$$S_{Nn} = \eta N^{-\frac{1}{2}}[\sigma_0^{-2}(X_{Nn}-a_0)]g_0(X_{Nn})/g_0(X_{Nn}) =$$

$$= \sigma_0^{-2}\,\eta N^{-\frac{1}{2}}(X_{Nn}-a_0) \quad \text{,voor } n = 1,2,\ldots,N \text{ en } N = 1,2,\ldots,$$

zodat

$$S_N = \sum_{n=1}^{N} S_{Nn} = \sigma_0^{-2}\eta N^{-\frac{1}{2}} \sum_{n=1}^{N} (X_{Nn}-a_0), \text{ voor } N = 1,2,\ldots .$$

De toets gebaseerd op S_N (equivalent met de Gauss-toets) is volgens stelling 2.5 een asymptotisch meest onderscheidende toets.

2.4. We kiezen voor μ de Lebesgue-maat op $(\mathbb{R},\mathcal{B})$ en $X_0 = \mathbb{R}$. Nu geldt

$$f_0(x) = \tfrac{1}{2}\exp(-|x|) > 0 \quad \text{op} \quad X_0,$$

en

$$\frac{\partial}{\partial\zeta} f_\zeta(x) = \begin{cases} f_\zeta(x) & \text{als} \quad x \geq \zeta, \\ -f_\zeta(x) & \text{als} \quad x < \zeta, \end{cases}$$

zodat $\dot{f}_0(x)$ bestaat voor alle $x \in X_0$.
Vervolgens verifieren we (2.18):

$$\int \dot{f}_0(x)dx = \int_{-\infty}^{0} -f_0(x)dx + \int_0^{\infty} f_0(x)dx = 0,$$

en

$$\int (\dot{f}_0^2(x)/f_0(x))dx = \int f_0(x)dx = 1 = J_0.$$

Ook aan (2.19) is voldaan daar

$$\int (\sqrt{f_\zeta(x)} - \sqrt{f_0(x)})^2 \cdot \zeta^{-2}dx = \frac{2}{\zeta^2}[1 - \int \sqrt{f_\zeta(x)}\,\sqrt{f_0(x)}\,dx] =$$

$$= \frac{2}{\zeta^2}[1-\tfrac{1}{2}\int \exp(-\tfrac{1}{2}|x|-\tfrac{1}{2}|x-\zeta|)dx] =$$

$$= 2\zeta^{-2}(1-(1+\tfrac{1}{2}|\zeta|)e^{-\frac{1}{2}|\zeta|}) \to \tfrac{1}{4}, \qquad \text{als } \zeta \to 0.$$

M.b.v. stelling 2.5 vinden we een asymptotisch meest onderscheidende toets gebaseerd op

$$S_N = \sum_{n=1}^{N} \frac{\eta}{N^{\frac{1}{2}}} \frac{\dot{f}_0(X_{Nn})}{f_0(X_{Nn})} = \eta N^{-\frac{1}{2}}(Y_N - Z_N),$$

waarbij Y_N gedefinieerd is als het aantal $X_{Nn} \geq 0$, en Z_N als het aantal $X_{Nn} < 0$, $n = 1,2,\ldots,N$, $N = 1,2,\ldots$.
We vinden dan de tekentoets terug als een asymptotisch meest onderscheidende toets.

2.5. Voor iedere ζ is $f_\zeta(x) > 0$ voor alle $x \in \mathbb{R}$ en geldt

$$\dot{f}_\zeta(x) := \partial f_\zeta(x)/\partial\zeta = \frac{2(x-\zeta)}{\pi[1+(x-\zeta)^2]^2},$$

voor alle x, zodat we in de notatie van voorwaarde (I) $X_0 = \mathbb{R}$ mogen kiezen. Voor μ kiezen we de Lebesgue-maat op $\mathbb{R}$.
Uit de uitdrukking voor $\dot{f}_\zeta$ zien we vervolgens dat

$$\int_{-\infty}^{\infty} \dot{f}_0(x)dx = \int_{-\infty}^{\infty} \frac{2x}{\pi(1+x^2)^2}\,dx = 0,$$

omdat de integraal bestaat en de integrand een oneven functie is.
Bovendien zien we dat

$$0 < J_0 = \int_{-\infty}^{\infty} [\dot{f}_0^2(x)/f_0(x)]dx = \int_{-\infty}^{\infty} \frac{4x^2}{\pi(1+x^2)^3}\,dx = \tfrac{1}{2} < \infty.$$

Aan (2.18) is dus blijkbaar voldaan.
Merk ten aanzien van (2.19) allereerst op dat op grond van de middelwaardestelling voor zekere $\tilde{\zeta} = \tilde{\zeta}(x)$, met $0 < |\tilde{\zeta}(x)| < \zeta$, geldt dat

$$[f_\zeta^{\frac{1}{2}}(x)-f_0^{\frac{1}{2}}(x)]^2\zeta^{-2} = \frac{(x-\tilde{\zeta})^2}{\pi[1+(x-\tilde{\zeta})^2]^3} \leq \frac{(|x|+1)^2}{\pi[1+x^2-|x|]^3},$$

voor alle x. We nemen hier aan dat ζ voldoet aan $|\zeta| < \frac{1}{2}$. Uit het bovenstaande volgt enerzijds dat $[f_\zeta^{\frac{1}{2}}(x)-f_0^{\frac{1}{2}}(x)]^2\zeta^{-2} \to \frac{1}{4}[\dot{f}_0^2(x)/f_0(x)]$ als $\zeta \to 0$, voor iedere x, terwijl anderzijds de uitdrukking begrensd wordt door de integreerbare functie $\pi^{-1}(|x|+1)^2[1+x^2-|x|]^{-3}$ mits $0 < |\zeta| < \frac{1}{2}$. Toepassing van de gedomineerde convergentiestelling van Lebesgue levert tenslotte dat "limiet" en "integratie" verwisseld mogen worden, zodat ook (2.19) geldt.
Kies vervolgens bijvoorbeeld $\theta(N,n) = 0$ en $\tau(N,n) = \eta N^{-\frac{1}{2}}$ voor $n = 1,2,\ldots,N$ en $N = 1,2,\ldots$ $(0 < \eta < \infty)$. Aan alle voorwaarden van stelling 2.5 is nu voldaan en toepassing van deze stelling geeft dat een AMP toets gebaseerd kan worden op

$$S_N := \eta N^{-\frac{1}{2}} \sum_{n=1}^{N} \frac{2X_{Nn}}{1+X_{Nn}^2},$$

$N = 1,2,\ldots$. Een AMP rij toetsen voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ is bijvoorbeeld de rij $\{\phi_N\} = \{\chi(\{S_N > 2^{-\frac{1}{2}}\eta\xi_{1-\alpha}\})\}$. Het asymptotisch onderscheidingsvermogen van deze rij toetsen is $1 - \Phi(\xi_{1-\alpha}-2^{-\frac{1}{2}}\eta)$.

2.6. a) De grootheid $\delta^2(P_{\theta(N,n)},P_{\tau(N,n)})$ is onafhankelijk van n en gelijk aan

$$\delta^2(P_{\theta(N,n)},P_{\tau(N,n)}) = \int_0^1 (\chi_{(0,1)}(x)-(1-\tau_N)^{-\frac{1}{2}}\chi_{(\tau_N,1)}(x))^2dx =$$

$$= 2[1-(1-\tau_N)^{\frac{1}{2}}],$$

zodat

$$\max_{1\le n\le N}[\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2 = 2[1-(1-\tau_N)^{\frac{1}{2}}],$$

$$\sum_{n=1}^{N}[\delta(P_{\theta(N,n)},P_{\tau(N,n)})]^2 = 2N[1-(1-\tau_N)^{\frac{1}{2}}].$$

Merk op dat zowel $\lim_{N\to\infty}2N[1-(1-\tau_N)^{\frac{1}{2}}] = a^2$ als $\tau_N = N^{-1}a^2+o(N^{-1})$, $N \to \infty$, impliceert dat $\lim_{N\to\infty}\tau_N = 0$, en dat dit laatste equivalent is met $\lim_{N\to\infty}2[1-(1-\tau_N)^{\frac{1}{2}}] = 0$. Het is dus voldoende om aan te tonen dat $\lim_{N\to\infty}2N[1-(1-\tau_N)^{\frac{1}{2}}] = a^2 \Leftrightarrow \tau_N = N^{-1}a^2+o(N^{-1})$, $N \to \infty$, onder de voor-

waarde dat $\lim_{N\to\infty}\tau_N = 0$. Onder deze voorwaarde geldt echter dat $2N[1-(1-\tau_N)^{\frac{1}{2}}] = N\tau_N + O(N\tau_N^2)$, $N \to \infty$, waaruit de te bewijzen equivalentie onmiddellijk volgt.

b) Alle kansmaten uit de familie worden gedomineerd door de restrictie van de Lebesgue-maat op $(0,1)$. Voor iedere $\zeta \in [0,1)$ is

$$\dot{f}_\zeta(x) := \partial f_\zeta(x)/\partial\zeta = \begin{cases} (1-\zeta)^{-2} & \text{voor } x \in (\zeta,1), \\ 0 & \text{voor } x \in (\zeta,1)^c. \end{cases}$$

De uitdrukking is niet gedefinieerd voor $x = \zeta$ en $x = 1$. I.h.b. is dus $f_0 > 0$ op $(0,1)$ en $\dot{f}_0 = 1$ op $(0,1)$, zodat we in de notatie van voorwaarde (I) $X_0 = (0,1)$ mogen kiezen.

Voorwaarde (2.18) is echter niet vervuld aangezien $\int_0^1 \dot{f}_0(x)dx = 1 > 0$; wèl geldt $0 < \int_0^1[\dot{f}_0^2(x)/f_0(x)]dx = 1 = J_0 < \infty$.

Evenmin is echter aan voorwaarde (2.19) voldaan, aangezien

$$\int_0^1 (\chi_{(0,1)}(x)-(1-\zeta)^{-\frac{1}{2}}\chi_{(\zeta,1)}(x))^2\zeta^{-2}dx =$$

$$= 2[1-(1-\zeta)^{\frac{1}{2}}]\zeta^{-2} \to \infty,$$

als $\zeta \to 0$. Deze limiet is ongelijk aan $J_0/4$ $(=\frac{1}{4})$.

Opmerking: Indien men de familie $\{P_\zeta, \zeta \geq 1\}$ van homogene kansverdelingen op $(0,\zeta)$ beschouwt en $\theta(N,n) = \theta_0 = 1$ en $\tau(N,n) = \tau_N > 1$ kiest voor $n = 1,2,\ldots,N$, $N = 1,2,\ldots$, kan men zelfs geen verzameling X_0 aanwijzen met de in voorwaarde (I) genoemde eigenschappen.

2.7. a) Zij $\lim_{N\to\infty} \int \phi_N \, dP_{\tau(N)} = \alpha$. We voeren het bewijs in twee stappen.

a1) Veronderstel eerst dat $\{P_{\tau(N)}\} \not\asymp \{P_{\theta(N)}\}$. Dan is er een rij verzamelingen $\{A_N\}$, met $A_N \in \mathcal{A}^N$, zodat $\lim_{N\to\infty} P_{\theta(N)}(A_N) = 0$ en $\limsup_{N\to\infty} P_{\tau(N)}(A_N) = a$, voor zekere $a > 0$. Kies $\{\psi_N\} = \{\chi(A_N) + \alpha\chi(A_N^c)\}$. Dan geldt dat

$$\int \psi_N \, dP_{\theta(N)} = P_{\theta(N)}(A_N) + \alpha(1-P_{\theta(N)}(A_N)) \to \alpha,$$

als $N \to \infty$, zodat $\{\psi_N\}$ een rij toetsen voor het asymptotische toetsings-

probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ is. Bovendien is

$$\limsup_{N\to\infty} \int \psi_N \, dP_{\tau(N)} =$$

$$= \limsup_{N\to\infty} [P_{\tau(N)}(A_N)+\alpha(1-P_{\tau(N)}(A_N))] = (1-\alpha)a+\alpha.$$

Maar hieruit volgt dat

$$\liminf_{N\to\infty} \int (\phi_N-\psi_N) dP_{\tau(N)} =$$

$$= \lim_{N\to\infty} \int \phi_N \, dP_{\tau(N)} - \limsup_{N\to\infty} \int \psi_N \, dP_{\tau(N)} = \alpha - (1-\alpha)a-\alpha < 0,$$

in strijd met het gegeven. Derhalve is $\{P_{\tau(N)}\} \lhd \{P_{\theta(N)}\}$.

a2) Veronderstel vervolgens dat $\{P_{\theta(N)}\} \not\lhd \{P_{\tau(N)}\}$. Dan is er een rij verzamelingen $\{B_N\}$, met $B_N \in \mathcal{A}^N$, zodat $\lim_{N\to\infty} P_{\tau(N)}(B_N) = 0$ en $\limsup_{N\to\infty} P_{\theta(N)}(B_N) = b$ voor zekere $b > 0$. Zij $b_N = P_{\theta(N)}(B_N)$ voor alle N en kies $\{\eta_N\} = \{[\alpha+\alpha(1-\alpha)b_N]\chi(B_N^c)\}$. Dan is, voor alle N,

$$\limsup_{N\to\infty} \int \eta_N \, dP_{\theta(N)} =$$

$$= \limsup_{N\to\infty} [\alpha+\alpha(1-\alpha)b_N](1-P_{\tau(N)}(B_N)) = \alpha+\alpha(1-\alpha)b.$$

Maar hieruit volgt dat

$$\liminf_{N\to\infty} \int (\phi_N-\eta_N) dP_{\tau(N)} =$$

$$= \lim_{N\to\infty} \int \phi_N \, dP_{\tau(N)} - \limsup_{N\to\infty} \int \eta_N \, dP_{\tau(N)} = \alpha-\alpha-\alpha(1-\alpha)b < 0,$$

in strijd met het gegeven. Daarom is $\{P_{\theta(N)}\} \lhd \{P_{\tau(N)}\}$.

Men kan het bewijs van a2) ook terugbrengen op a1). Zij ϕ_N^* de meest onderscheidende toets bij het probleem $(\alpha;\theta(N),\tau(N))$. Dan geldt uiteraard ook $\lim_{N\to\infty} \int \phi_N^* dP_{\tau(N)} = \alpha$. Uit het lemma van Neyman-Pearson leidt men dan direct af dat $1 - \phi_N^*$ een meest onderscheidende toets is van het probleem $(1-\int \phi_N^* \, dP_{\tau(N)}; \tau(N),\theta(N))$, met onderscheidingsvermogen $1 - \int \phi_N^* \, dP_{\theta(N)}$.

Uit het voorgaande volgt dat $\lim_{N\to\infty}(1-\int\phi_N^* \, dP_{\theta(N)}) = 1-\alpha$.
Nu is voldaan aan de gegevens bij a1), met $1-\alpha$ i.p.v. α en met verwisseling van de rollen van $\theta(N)$ en $\tau(N)$.
De in a1) bewezen implicatie levert dus nu $\{P_{\theta(N)}\} \lhd \{P_{\tau(N)}\}$.

b) Zij $\lim_{N\to\infty} \int\phi_N \, dP_{\tau(N)} = 1$. Onderstel dat $\{P_{\theta(N)}\} \lhd \{P_{\tau(N)}\}$.
Dan gelden de implicaties:

$$\lim_{N\to\infty} \int\phi_N \, dP_{\tau(N)} = 1 \Rightarrow \phi_N \xrightarrow{P}_{\tau(N)} 1 \Rightarrow \phi_N \xrightarrow{P}_{\theta(N)} 1 \Rightarrow \lim_{N\to\infty} \int\phi_N \, dP_{\theta(N)} = 1,$$

in strijd met het gegeven dat $\{\phi_N\}$ asymptotische onbetrouwbaarheidsdrempel $\alpha < 1$ heeft. Dus is $\{P_{\theta(N)}\} \not\lhd \{P_{\tau(N)}\}$.

Opmerking 1: Als in de opgave $\alpha = 0$ gekozen wordt, blijft (b) doorgaan en kunnen we aan de conclusie nog toevoegen dat $\{P_{\tau(N)}\} \not\lhd \{P_{\theta(N)}\}$. In (a) kunnen we dan slechts concluderen dat $\{P_{\tau(N)}\} \lhd \{P_{\theta(N)}\}$.

Opmerking 2: Het omgekeerde van de bewering in a) is i.h.a. onjuist. Zie bijv. stelling 2.5.

2.8. De tekentoets is gebaseerd op de stochastische grootheid Z_N, gedefinieerd als het aantal der $X_{N1},\ldots,X_{NN}$ dat ≥ 0 is. Onder de nulhypothese heeft Z_N een $\mathcal{B}i(N;\frac{1}{2})$ verdeling, zodat $2(Z_N-\frac{1}{2}N)N^{-\frac{1}{2}}$ asymptotisch een standaardnormale verdeling bezit. Verwerp de nulhypothese als $Z_N \geq$ $\geq \frac{1}{2}(N+N^{\frac{1}{2}}\xi_{1-\alpha})$. Op grond van de zojuist genoemde asymptotische normaliteit heeft deze toets een asymptotische onbetrouwbaarheid α.
Voor de berekening van het asymptotische onderscheidingsvermogen merken we op dat Z_N onder het alternatief een $\mathcal{B}i(N;p_N)$ verdeling heeft, waarin $p_N = \Phi(\eta N^{-\frac{1}{2}})$, zodat $(Z_N-Np_N)\,[Np_N(1-p_N)]^{-\frac{1}{2}}$ asymptotisch een standaardnormale verdeling bezit. Hieruit volgt dat

$$P_{\tau(N)}(Z_N \geq \tfrac{1}{2}(N+N^{\frac{1}{2}}\xi_{1-\alpha})) =$$

$$= P_{\tau(N)}((Z_N-Np_N)[Np_N(1-p_N)]^{-\frac{1}{2}} \geq [\tfrac{1}{2}(N+N^{\frac{1}{2}}\xi_{1-\alpha})-Np_N][Np_N(1-p_N)]^{-\frac{1}{2}}) =$$

$$= P_{\tau(N)}((Z_N-Np_N)[Np_N(1-p_N)]^{-\frac{1}{2}} \geq [\tfrac{1}{2}N^{\frac{1}{2}}\xi_{1-\alpha}+\tfrac{1}{2}N-N\Phi(0) -$$

$$- \eta N^{\frac{1}{2}}\Phi'(0)+\mathcal{O}(1)][Np_N(1-p_N)]^{-\frac{1}{2}}) \to 1 - \Phi(\xi_{1-\alpha}-\eta(2/\pi)^{\frac{1}{2}}).$$

De Gauss-toets is gebaseerd op $S_N := N^{-\frac{1}{2}} \sum_{n=1}^{N} X_{Nn}$. Onder de nulhypothese heeft S_N een $N(0;1)$ verdeling voor alle N. Men verwerpt de nulhypothese als $S_N \geq \xi_{1-\alpha}$. Voor iedere N is de onbetrouwbaarheid van deze toets gelijk aan α.
Onder het alternatief bezit S_N voor alle N een $N(\eta;1)$ verdeling, zodat

$$P_{\tau(N)}(S_N \geq \xi_{1-\alpha}) = P_{\tau(N)}(S_N - \eta \geq \xi_{1-\alpha} - \eta) = 1 - \Phi(\xi_{1-\alpha} - \eta),$$

voor alle N.
Uit definitie 2.3 volgt nu dat de ARE van de tekentoets ten opzichte van de Gauss-toets gelijk is aan $2/\pi$.

2.9. Definieer $\bar{X}_{N,1} = N_1^{-1} \sum_{n=1}^{N_1} X_{Nn}$ en $\bar{X}_{N,2} = N_2^{-1} \sum_{n=N_1+1}^{N} X_{Nn}$. De tweesteekproeventoets van Student is gebaseerd op

$$T_N := N^{-\frac{1}{2}} N_1^{\frac{1}{2}} N_2^{\frac{1}{2}} (\bar{X}_{N,1} - \bar{X}_{N,2}) \left[\frac{\sum_{n=1}^{N_1} (X_{Nn} - \bar{X}_{N,1})^2 + \sum_{n=N_1+1}^{N} (X_{Nn} - \bar{X}_{N,2})^2}{N-2} \right]^{-\frac{1}{2}}.$$

Het is bekend dat

$$(N_1 - 1)^{-1} \sum_{n=1}^{N_1} (X_{Nn} - \bar{X}_{N,1})^2 = \sigma^2 + R_{N,1},$$

$$(N_2 - 1)^{-1} \sum_{n=N_1+1}^{N} (X_{Nn} - \bar{X}_{N,2})^2 = \sigma^2 + R_{N,2},$$

waarin, voor $N \to \infty$,

$$R_{N,i} \to_{P_{\theta(N)}} 0, \quad R_{N,i} \to_{P_{\tau(N)}} 0, \qquad i = 1,2.$$

Hieruit volgt dat de term tussen rechte haken in de uitdrukking voor T_N zowel onder de nulhypothese als het alternatief in kans naar σ^2 convergeert.
Onder de nulhypothese heeft T_N een centrale Student-verdeling met N-2 vrijheidsgraden. Op grond van het bovenstaande en lemma 1.8 heeft T_N asymptotisch een $N(0;1)$ verdeling. Verwerp de nulhypothese als $T_N \geq \xi_{1-\alpha}$. De asymptotische onbetrouwbaarheid van deze toets is dus α.

Op analoge manier vinden we met behulp van lemma 1.8 dat T_N onder het alternatief asymptotisch een $N(\eta\sigma^{-1};1)$ verdeling bezit, zodat

$$P_{\tau(N)}(T_N \geq \xi_{1-\alpha}) = P_{\tau(N)}(T_N - \eta\sigma^{-1} \geq \xi_{1-\alpha} - \eta\sigma^{-1}) \to 1 - \Phi(\xi_{1-\alpha} - \eta\sigma^{-1}),$$

als $N \to \infty$.

De twee-steekproeven toets van Gauss is gebaseerd op

$$S_N := N^{-\frac{1}{2}} N_1^{\frac{1}{2}} N_2^{\frac{1}{2}} (\bar{X}_{N,1} - \bar{X}_{N,2})\sigma^{-1}.$$

Onder de nulhypothese heeft S_N een $N(0;1)$ verdeling voor alle N.
Men verwerpt de nulhypothese als $S_N \geq \xi_{1-\alpha}$. Voor iedere N is de onbetrouwbaarheid van deze toets gelijk aan α.
Onder het alternatief bezit S_N een $N(\eta\sigma^{-1};1)$ verdeling voor alle N, zodat

$$P_{\tau(N)}(S_N \geq \xi_{1-\alpha}) = P_{\tau(N)}(S_N - \eta\sigma^{-1} \geq \xi_{1-\alpha} - \eta\sigma^{-1}) = 1 - \Phi(\xi_{1-\alpha} - \eta\sigma^{-1}),$$

voor alle N.
Uit definitie 2.3 volgt dat de ARE van de twee-steekproeventoets van Student ten opzichte van de twee-steekproeventoets van Gauss gelijk is aan 1.

Vraagstukken

3.1. In het algemeen zullen de scores $a_N(n,\gamma_0)$, $n = 1,2,\ldots,N$, niet alleen van de familie verdelingsfuncties $\{G_\zeta, \zeta \in T\}$ maar ook nog van het inwendige punt $\theta_0 \in T$ afhangen. Deze afhankelijkheid van θ_0 is echter niet aanwezig bij het lokatiemodel (d.w.z. $G_\zeta(x) = G(x-\zeta)$, $-\infty < x < \infty$, $\zeta \in T$, waarin G een vaste verdelingsfunctie is).
In dit geval vinden we dat

$$\gamma_0(t) = \frac{\dot{g}_0(G_0^{-1}(t))}{g_0(G_0^{-1}(t))} = -\frac{g'(G^{-1}(t))}{g(G^{-1}(t))}, \qquad 0 < t < 1,$$

waarbij g' de afgeleide (naar x) van g voorstelt (zie opmerking 3.4, hoofdstuk 3, pag. 75).
Bereken deze scorefunctie in de volgende gevallen:
a) G is de standaardnormale verdelingsfunctie,
b) G is de verdelingsfunctie van een normale verdeling met variantie σ^2,
c) G is de standaardlogistieke verdelingsfunctie,
d) G is de dubbel-exponentiële verdelingsfunctie.

3.2. Ga na dat de symmetrie-toets van Wilcoxon de lokaal meest onderscheidende rangtoets is voor de hypothese H_1 (symmetrie) tegen logistieke verschuivingsalternatieven ($G(x) = (1+e^{-x})^{-1}$, $-\infty < x < \infty$).

3.3. Zij $\{G_\zeta, \zeta \in \mathbb{R}\}$ de familie van Cauchy-verdelingen op $(\mathbb{R},\mathcal{B})$ met dichtheden

$$g_\zeta(x) = \frac{1}{\pi} \frac{1}{1+(x-\zeta+\theta_0)^2},$$

t.o.v. de Lebesgue-maat op $\mathbb{R}$ (vergelijk opgave 2.5).
Laat zien dat deze familie dichtheden aan voorwaarde (II) (hoofdstuk 3, pag. 72) voldoet.
Construeer m.b.v. stelling 3.7 de lokaal meest onderscheidende rangtoets voor het toetsingsprobleem $(\alpha;H_0,K_c)$, $0 < \alpha < 1$.

3.4. Zij $\{G_\zeta, \zeta \in T \subset \mathbb{R}\}$ de familie dubbel-exponentiële verdelingen op $(\mathbb{R},\mathcal{B})$ met dichtheden

$$g_\zeta(x) = \tfrac{1}{2} \exp(-|x-\zeta+\theta_0|),$$

t.o.v. de Lebesgue-maat op $\mathbb{R}$ (vergelijk opgave 2.4).
Laat zien dat deze familie dichtheden aan voorwaarde (II) (hoofdstuk 3, pag. 72) voldoet.
Construeer m.b.v. stelling 3.7 de lokaal meest onderscheidende rang-toets voor het toetsingsprobleem $(\alpha;H_0,K_c)$, $0 < \alpha < 1$, in het speciale geval der twee-steekproef situatie.

3.5. Een k-steekproeven toets tegen trend.
Zij $\{G_\zeta, \zeta \in T \subset \mathbb{R}\}$ de familie logistieke verdelingen met dichtheden

$$g_\zeta(x) = \frac{e^{-x-\zeta+\theta_0}}{(1+e^{-x-\zeta+\theta_0})^2},$$

t.o.v. de Lebesgue-maat op $\mathbb{R}$. Zij $0 < \alpha < 1$.
Laat zien m.b.v. stelling 3.13 (voor deze klasse van verdelingen) dat de toets die verwerpt voor "grote" waarden van

$$\sum_{l=1}^{k} l \sum_{n=N_{l-1}+1}^{N_l} R_{Nn},$$

asymptotisch uniform meest onderscheidend is voor het asymptotische toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$, als $c_{Nn} = l$ voor $n = N_{l-1}+1,\ldots,N_l$ $(l=1,2,\ldots,k)$, $N_1 < N_2 < \ldots < N_k = N$ en Δ_N zodanig dat

$$\lim_{N\to\infty} \Delta_N^2 \max_{1\le n\le N} (c_{Nn}-\bar{c}_N)^2 = 0$$

én

$$\lim_{N\to\infty} \Delta_N^2 \sum_{n=1}^{N} (c_{Nn}-\bar{c}_N)^2 \to \eta^2,$$

voor zekere η, $0 < \eta < \infty$.

3.6. Bereken voor het symmetrieprobleem de ARE van de toets van Wilcoxon t.o.v. de tekentoets in het geval van verschuivingsalternatieven van logistieke verdelingen, waarbij de verschuivingen van de orde $N^{-\frac{1}{2}}$ zijn.

3.7. Zij $\{G_\zeta, \zeta \in T \subset \mathbb{R}\}$ de familie van dubbel-exponentiële verdelingen op $(\mathbb{R},\mathcal{B})$ met dichtheden

$$g_\zeta(x) = \tfrac{1}{2} \exp(-|x-\zeta+\theta_0|),$$

t.o.v. de Lebesgue-maat op $\mathbb{R}$. Vergelijk opgave 3.4.
Construeer een asymptotisch uniform meest onderscheidende toets voor het asymptotische toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$, $0 < \alpha < 1$, in het speciale geval der twee-steekproef situatie, gebruikmakend van benaderende scores.

3.8. Bereken m.b.v. stelling 3.18 de ARE van de toets van Fisher-Yates t.o.v. de toets van Wilcoxon in het twee-steekproeven probleem (bij naar oneindig convergerende steekproefomvangen), voor het toetsen van $\{H_{0N}\}$ tegen

a) rijen normale verschuivingsalternatieven,
b) rijen logistieke verschuivingsalternatieven,

welke voldoen aan (3.43) en (3.44) voor zekere $\eta > 0$, $0 < \alpha < 1$.

3.9. Bereken m.b.v. stelling 3.18 de ARE van de toets van Wilcoxon t.o.v. de mediaan-toets (vergelijk opgave 3.7) in het twee-steekproeven probleem (bij naar oneindig convergerende steekproefomvangen), voor het toetsen van $\{H_{0N}\}$ tegen rijen dubbel-exponentiële verschuivingsalternatieven welke voldoen aan (3.43) en (3.44) voor zekere $\eta > 0$, $0 < \alpha < 1$.

3.10: Zij $\theta(N,n) = \theta_0$ en $\tau(N,n) = \tau_N(\eta) = \theta_0 + \eta N^{-\frac{1}{2}} + o(N^{-\frac{1}{2}})$ voor alle $N = 1,2,\ldots$, $n = 1,2,\ldots,N$ en willekeurige $\eta > 0$ $(N \to \infty)$.
Zij $0 < \alpha < 1$ en laten $\{\psi_{1N}\}$ en $\{\psi_{2N}\}$ twee rijen toetsen zijn voor het probleem $(\alpha;\{\theta(N)\},\{\tau(N)\})$ met de eigenschap dat

$$\lim_{N\to\infty} \int \psi_{iN}\, dP_{\tau(N)} = \lim_{N\to\infty} \int \psi_{iN}\, dP^N_{\tau_N(\eta)} =$$

$$= 1 - \Phi(\xi_{1-\alpha} - \eta\delta_i), \qquad i = 1,2,$$

voor zekere $\delta_i > 0$ $(i=1,2)$ en alle $\eta > 0$.
a) Bewijs dat

$$\lim_{N\to\infty} \int \psi_{i\tilde{N}(N)}\, dP^{\tilde{N}(N)}_{\tau_N(\eta)} = 1 - \Phi(\xi_{1-\alpha} - c\eta\delta_i), \qquad i = 1,2,$$

voor iedere rij natuurlijke getallen $\{\tilde{N}(N)\}$, waarvoor $\lim_{N\to\infty}[\tilde{N}(N)]^{\frac{1}{2}}N^{-\frac{1}{2}} =$
$= c > 0$.
b) Toon aan dat er een rij natuurlijke getallen $\{\tilde{N}\} = \{\tilde{N}(N)\}$ bestaat zodanig dat

$$\lim_{N\to\infty} \int \psi_{2\tilde{N}(N)}\, dP^{\tilde{N}(N)}_{\tau_N(\eta)} = \lim_{N\to\infty} \int \psi_{1N}\, dP^N_{\tau_N(\eta)},$$

en

$$\lim_{N\to\infty} \frac{\tilde{N}(N)}{N} = \left(\frac{\delta_1}{\delta_2}\right)^2 = \mathrm{ARE}(\alpha;\{\psi_{1N}\},\{\psi_{2N}\};\{\tau(N)\}).$$

Oplossingen der vraagstukken

3.1. a) Bij een standaardnormale verdelingsfunctie is $g(x) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}x^2)$, zodat $g'(x) = -xg(x)$. Voor $0 < t < 1$ geldt nu

$$\gamma_0(t) = \frac{-g'(G^{-1}(t))}{g(G^{-1}(t))} = \frac{\Phi^{-1}(t)g(\Phi^{-1}(t))}{g(\Phi^{-1}(t))} = \Phi^{-1}(t).$$

b) Vanwege de onafhankelijkheid der scorefunctie van de lokatie-parameter kunnen we uitgaan van een normale verdelingsfunctie met verwachting 0 en variantie σ^2.
Dan is $g(x) = (2\pi)^{-\frac{1}{2}}\sigma^{-1} \exp(-x^2/2\sigma^2)$, zodat $g'(x) = -(x/\sigma^2)g(x)$.
Voor $0 < t < 1$ geldt nu

$$\gamma_0(t) = \frac{-g'(G^{-1}(t))}{g(G^{-1}(t))} = \frac{\sigma^{-2}G^{-1}(t)g(G^{-1}(t))}{g(G^{-1}(t))} = \sigma^{-2}G^{-1}(t) = \sigma^{-1}\Phi^{-1}(t),$$

waarbij $G(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}}\sigma^{-1} \exp(-y^2/2\sigma^2)dy$.

c) Bij een standaardlogistieke verdelingsfunctie is $G(x) = (1+e^{-x})^{-1}$, $g(x) = e^{-x}G^2(x)$ zodat $g'(x) = e^{-x}G^2(x)(2G(x)e^{-x}-1) =$
$= e^{-x}G^2(x)(1-2G(x))$. Voor $0 < t < 1$ geldt

$$\gamma_0(t) = \frac{-g'(G^{-1}(t))}{g(G^{-1}(t))} = 2t - 1.$$

d) Bij de dubbel-exponentiële verdelingsfunctie is $g(x) = \frac{1}{2}\exp(-|x|)$, $-\infty < x < \infty$, zodat $g'(x) = -\frac{1}{2}\exp(-x)$ als $x > 0$ en $g'(x) = \frac{1}{2}\exp(x)$ als $x < 0$, en dus $-g'(x)/g(x) = 1$ als $x > 0$ en $-g'(x)/g(x) = -1$ als $x < 0$.
Voor $0 < t < 1$ geldt $\gamma_0(t) = 1$ als $G^{-1}(t) > 0$ en $\gamma_0(t) = -1$ als $G^{-1}(t) < 0$, ofwel $\gamma_0(t) = \text{sgn}(2t-1)$.

3.2. We passen stelling 3.9 toe. In opgave 3.1 is reeds berekend dat $\gamma_0(t) = 2t - 1$ bij logistieke verdelingen, zodat $\gamma_0^+(t) = \gamma_0(\frac{1}{2}+\frac{1}{2}t) = 2(\frac{1}{2}+\frac{1}{2}t) - 1 = t$, $0 < t < 1$. De scores behorende bij deze scorefunctie γ_0^+ zijn

$$a_N(n,\gamma_0^+) = E\gamma_0^+(U^{(n)}) = EU^{(n)} = n(N+1)^{-1}, \text{ voor } n = 1,2,\ldots,N.$$

De lokaal meest onderscheidende rangtoets voor het toetsen van H_1 tegen logistieke verschuivingsalternatieven (gebaseerd op de aselecte steekproef $X_1,\ldots,X_N$) verwerpt derhalve voor "grote" waarden van

$$\frac{1}{N+1}\sum_{n=1}^{N} R_n^+ \operatorname{sgn} X_n.$$

Deze toets is equivalent met de symmetrietoets van Wilcoxon.

3.3. Voor alle $x \in \mathbb{R}$ geldt $\dot{g}_0(x) = [\partial g_\zeta(x)/\partial\zeta]_{\zeta=\theta_0} = 2x/(\pi(1+x^2)^2)$, zodat we in de notatie van voorwaarde (II) $N_0 = \emptyset$ mogen kiezen en voor μ de Lebesgue-maat op $\mathbb{R}$. Hieruit volgt

$$\int_{-\infty}^{\infty} |\dot{g}_0(x)|\,d\mu(x) = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{|x|}{(1+x^2)^2}\,dx = \frac{2}{\pi} < \infty.$$

Merk ten aanzien van (3.9) allereerst op dat op grond van de middelwaardestelling voor zekere $\hat{\zeta} = \hat{\zeta}(x)$ geldt dat

$$\left|\frac{g_\zeta(x)-g_0(x)}{\zeta-\theta_0}\right| = \frac{2}{\pi}\left|\frac{(x-\hat{\zeta}+\theta_0)}{\{1+(x-\hat{\zeta}+\theta_0)^2\}^2}\right| \le \frac{2}{\pi}\frac{|x|+1}{\{1+x^2-|x|\}^2},$$

voor alle x, mits $|\zeta - \theta_0| < \frac{1}{2}$. Aangezien

$$\frac{2}{\pi}\int_{-\infty}^{\infty} \frac{|x|+1}{\{1+x^2-|x|\}^2}\,dx < \infty,$$

geeft toepassing van de gedomineerde convergentiestelling van Lebesgue het gewenste resultaat (3.9). Volgens stelling 3.7 verwerpt de lokaal meest onderscheidende rangtoets voor het toetsingsprobleem $(\alpha;H_0,K_c)$ voor "grote" waarden van $T = \sum_{n=1}^{N} c_n a_N(r_n,\gamma_0)$, waarbij $r_1,\ldots,r_N$ de waarden zijn der rangnummers van de steekproefelementen $X_1,\ldots,X_N$. Hierin is

$$\gamma_0(t) = \dot{g}_0(G_0^{-1}(t))/g_0(G_0^{-1}(t)), \qquad \text{voor } 0 < t < 1.$$

Daar we in opmerking 3.4 gezien hebben dat de scorefunctie niet van θ_0 afhangt, kiezen we $\theta_0 = 0$. Dan is $\dot{g}_0(x) = 2x/(\pi(1+x^2)^2)$ en $g_0(x) = (\pi(1+x^2))^{-1}$ zodat $\dot{g}_0(x)/g_0(x) = 2x/(1+x^2)$ en dus

$$\gamma_0(t) = \frac{2G_0^{-1}(t)}{1+(G_0^{-1}(t))^2} .$$

Uit $G_0(x) = \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{1+y^2}\,dy$ kunnen we afleiden dat $G_0^{-1}(t) = tg(\pi(t-\frac{1}{2}))$. De scores $a_N(n,\gamma_0)$, $(n=1,2,\ldots,N)$, volgen op de bekende wijze uit de scorefunctie.

3.4. Allereerst geldt $g_0(x) = g_{\theta_0}(x) = \frac{1}{2}\exp(-|x|) > 0$ en

$$G_0(x) = \tfrac{1}{2}\int_{-\infty}^{x} \exp(-|y|)dy = 1 - \tfrac{1}{2}\exp(-x), \quad \text{voor alle } x \in \mathbb{R}.$$

Vervolgens merken we op dat $\dot{g}_0(x) = -\frac{1}{2}\exp(-x)$ als $x > 0$ en $\dot{g}_0(x) = \frac{1}{2}\exp(x)$ als $x < 0$, zodat

$$\int_{-\infty}^{\infty} |\dot{g}_0(x)|\,dx = \int_{-\infty}^{x} \tfrac{1}{2}\exp(x)dx + \int_{0}^{\infty} \tfrac{1}{2}\exp(-x)dx = 1 < \infty.$$

We kunnen kiezen $N_0 = \{0\}$, dan rest ons alleen nog te bewijzen dat

$$\lim_{\zeta\to\theta_0} \int_{-\infty}^{\infty} \left|\frac{g_\zeta(x)-g_{\theta_0}(x)}{\zeta-\theta_0}\right| dx = 1.$$

Neem aan dat $\zeta > \theta_0$ (het geval dat $\zeta < \theta_0$ verloopt net zo), dan is

$$\lim_{\zeta\to\theta_0} \int_{-\infty}^{\infty} \left|\frac{g_\theta(x)-g_{\theta_0}(x)}{\zeta-\theta_0}\right| dx =$$

$$= \lim_{\zeta\to\theta_0} \tfrac{1}{2}\int_{-\infty}^{\infty} \left|\frac{\exp(-|x-\zeta+\theta_0|)-\exp(-|x|)}{\zeta-\theta_0}\right| dx =$$

$$= \lim_{\zeta\to\theta_0} \tfrac{1}{2}\left\{\int_{-\infty}^{0} \left|\frac{\exp(x-\zeta+\theta_0)-\exp(x)}{\zeta-\theta_0}\right| dx + \right.$$

$$\left. + \int_{0}^{\zeta-\theta_0} \left|\frac{\exp(x-\zeta+\theta_0)-\exp(-x)}{\zeta-\theta_0}\right| dx + \int_{\zeta-\theta_0}^{\infty} \left|\frac{\exp(-x+\zeta-\theta_0)-\exp(-x)}{\zeta-\theta_0}\right| dx\right\} =$$

$$= \lim_{\zeta\to\theta_0} \tfrac{1}{2}\left\{\int_{-\infty}^{0} \frac{\exp(x)-\exp(x-\zeta+\theta_0)}{\zeta-\theta_0}\,dx + \int_{0}^{\frac{1}{2}(\zeta-\theta_0)} \frac{\exp(-x)-\exp(x-\zeta+\theta_0)}{\zeta-\theta_0}\,dx + \right.$$

$$\left. + \int_{\frac{1}{2}(\zeta-\theta_0)}^{\zeta-\theta_0} \frac{\exp(x-\zeta+\theta_0)-\exp(-x)}{\zeta-\theta_0}\,dx + \int_{\zeta-\theta_0}^{\infty} \frac{\exp(-x+\zeta-\theta_0)-\exp(-x)}{\zeta-\theta_0}\,dx \right\} =$$

$$= \lim_{\zeta\to\theta_0} (2-2\exp(-\tfrac{1}{2}(\zeta-\theta_0))) \to 1, \qquad \text{als } \zeta \to \theta_0,$$

zodat aan voorwaarde (II) is voldaan. In de twee-steekproef situatie geldt $c_n = 1$ als $n = 1,2,\ldots,N_1$ en $c_n = 0$ als $n = N_1+1,\ldots,N_1+N_2$, waarbij N_1 en N_2 de steekproefomvangen zijn van de eerste respectievelijk de tweede steekproef. Met behulp van stelling 3.7 vinden we dat de lokaal meest onderscheidende rangtoets voor het toetsingsprobleem $(\alpha;H_0,K_c)$ verwerpt voor "grote" waarden van

$$T = \sum_{n=1}^{N_1} c_n a_N(r_n,\gamma_0),$$

waarbij $r_1,\ldots,r_N$ de waarden zijn der rangnummers van $X_1,X_2,\ldots,X_{N_1}$, $X_{N_1+1},\ldots,X_{N_1+N_2}$ ($X_1,\ldots,X_{N_1}$ vormen de eerste steekproef en $X_{N_1+1},\ldots,X_{N_1+N_2}$ vormen de tweede steekproef). Bovendien is

$$\gamma_0(t) = \operatorname{sgn}(2t-1), \qquad 0 < t < 1 \quad \text{(zie opgave 3.1)},$$

zodat

$$a_N(n,\gamma_0) = E\ \operatorname{sgn}(2U^{(n)}-1),$$

waarbij $U^{(1)},\ldots,U^{(N)}$ een geordende steekproef voorstelt van omvang N uit een homogene verdeling op (0,1).

3.5. We verifiëren eerst voorwaarde (I) (pag. 90). Voor logistieke verdelingen is $\gamma_0(t) = 2t - 1$, $(0<t<1)$. Daar $\gamma_0(t) = -g_0'(G_0^{-1}(t))/g_0(G_0^{-1}(t))$ gaat men eenvoudig na dat

$$J_0 = \int \frac{g_0'^2}{g_0^2} g_0 d\lambda = \int_0^1 (2t-1)^2 dt = \frac{1}{3}, \text{ zodat } 0 < J_0 < \infty.$$

Relatie (3.29) volgt onmiddellijk uit de gedomineerde convergentiestelling van Lebesgue daar de integrand in het linkerlid van (3.29) voor alle $|\zeta-\theta_0| < 1$ gedomineerd wordt door de integreerbare functie

$$\left[\frac{\exp(-\frac{1}{2}x)(\exp(-x)+1)}{(1+\exp(-x+1))(1+\exp(-x))}\right]^2 .$$

Vervolgens onderzoeken we nader de gedaante van

$$\tilde{T}_N = J_0^{-\frac{1}{2}}\Big(\sum_{n=1}^{N} (c_{Nn}-\bar{c}_N)^2\Big)^{-\frac{1}{2}} \sum_{n=1}^{N} (c_{Nn}-\bar{c}_N) a_N(R_{Nn},\gamma_0).$$

Daar

$$a_N(n,\gamma_0) = E(2U^{(n)}-1) = 2n/(N+1)-1$$

is

$$\tilde{T}_N = J_0^{-\frac{1}{2}}\Big(\sum_{n=1}^{N} (c_{Nn}-\bar{c}_N)^2\Big)^{-\frac{1}{2}} \sum_{l=1}^{k} \sum_{n=N_{l-1}+1}^{N_l} (1-N^{-1}\sum_{l=1}^{k} lN_l)(2R_{Nn}(N+1)^{-1}-1).$$

Op grond van stelling 3.13 verwerpt de asymptotisch meest onderscheidende toets voor het toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$ voor "grote" waarden van $\tilde{T}_N$, hetgeen op hetzelfde neerkomt als verwerpen voor "grote" waarden van

$$\sum_{l=1}^{k} l \sum_{n=N_{l-1}}^{N_l} R_{Nn}.$$

3.6. De tekentoets is gebaseerd op de stochastische grootheid Z_N, gedefinieerd als het aantal der $X_{N1},\ldots,X_{NN}$ dat niet-negatief is. Onder de nulhypothese heeft Z_N een $\mathcal{B}i(N;\frac{1}{2})$ verdeling, zodat $2(Z_N-\frac{1}{2}N)N^{-\frac{1}{2}}$ asymptotisch een standaardnormale verdeling bezit. Verwerpen we de nulhypothese als $Z_N \geq \frac{1}{2}(N+N^{\frac{1}{2}}\xi_{1-\alpha})$ dan heeft deze toets op grond van de zojuist genoemde asymptotische normaliteit een asymptotische onbetrouwbaarheid α.

Voor de berekening van het asymptotische onderscheidingsvermogen merken we op dat Z_N onder het alternatief een $\mathcal{B}i(N,p_N)$ verdeling heeft, waarin voor vaste $\eta > 0$, $p_N = F(\eta N^{-\frac{1}{2}})$, met $F(x) = (1+\exp(-x))^{-1}$, $-\infty < x < \infty$, zodat $(Z_N - Np_N)\,[Np_N(1-p_N)]^{-\frac{1}{2}}$ asymptotisch een standaardnormale verdeling bezit. Hieruit volgt dat onder het verschuivingsalternatief $\Delta_N = \eta N^{-\frac{1}{2}}$

$$P(Z_N \geq \tfrac{1}{2}(N+N^{\frac{1}{2}}\xi_{1-\alpha})) = P((Z_N-Np_N)[Np_N(1-p_N)]^{-\frac{1}{2}} \geq$$

$$\geq [\tfrac{1}{2}(N+N^{\frac{1}{2}}\xi_{1-\alpha})-NP_N][Np_N(1-p_N)]^{-\frac{1}{2}}) = P((Z_N-Np_N)[Np_N(1-p_N)]^{-\frac{1}{2}} \geq$$

$$[\tfrac{1}{2}N^{\frac{1}{2}}\xi_{1-\alpha}+\tfrac{1}{2}N-NF(0)-\eta N^{\frac{1}{2}}F'(0)+o(1)][Np_N(1-p_N)]^{-\frac{1}{2}} \to 1-\Phi(\xi_{1-\alpha}-\tfrac{1}{2}\eta).$$

Om het asymptotische onderscheidingsvermogen van de symmetrietoets van Wilcoxon te berekenen passen we stelling 3.19 toe. In opgave 3.5 is reeds uitgewerkt dat voor logistieke verdelingen geldt $\gamma_0(t) = 2t - 1$, $(0<t<1)$, zodat $\gamma^+(t) = t$ en $J_0 = \frac{1}{3}$. Vervolgens merken we op dat we in opgave 3.2 reeds hebben gezien dat stelling 3.19 de symmetrietoets van Wilcoxon oplevert in deze situatie en dat het asymptotisch onderscheidingsvermogen gelijk is aan $1 - \Phi(\xi_{1-\alpha} - \frac{1}{3}\eta\sqrt{3})$, zie (3.74). Uit definitie (2.3) volgt nu dat de ARE van de tekentoets ten opzichte van de toets van Wilcoxon gelijk is aan $(\frac{1}{2}(\frac{1}{3}\sqrt{3})^{-1})^2 = \frac{3}{4}$.

3.7. We passen gevolg 3.14 in combinatie met stelling 3.16 toe. Dat de dichtheid g_0 in het enkele punt $x = 0$ niet differentieerbaar is, stoort niet. (Men gaat ook eenvoudig na dat voorwaarde (I) vervuld is).
Aan de voorwaarden van stelling 3.13 is voldaan en bovendien voldoen de benaderende scores $a_N(n) = \gamma_0(\frac{n}{N+1})$, $N = 1,2,\ldots$, $n = 1,2,\ldots,N$ op grond van lemma 3.17 aan (3.49), immers $\gamma_0(t) = \mathrm{sgn}(2t-1)$. Toepassing van stelling 3.16 levert ons bij alternatieven die aan (3.44) voldoen dat een asymptotisch uniform meest onderscheidende toets voor het asymptotische toetsingsprobleem $(\alpha;\{H_{0N}\},\{K_{cN}\})$, $0 < \alpha < 1$, verwerpt voor "grote" waarden van

$$\tilde{T}_N = J_0^{-\frac{1}{2}}\Big(\sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)^2\Big)^{-\frac{1}{2}}\sum_{n=1}^{N}(c_{Nn}-\bar{c}_N)\mathrm{sgn}(2R_{Nn}/(N+1)-1).$$

In het speciale geval van het twee-steekproeven probleem geldt dat $c_{Nn} = 1$ voor $n = 1,2,\ldots,N_1$, $c_{Nn} = 0$ voor $n = N_1+1,\ldots,N_1+N_2$, zodat we eenvoudig kunnen nagaan dat de aldus geconstrueerde toets equivalent is met de mediaantoets, die de nulhypothese verwerpt voor "grote" waarden van de toetsingsgrootheid T'_N : het aantal waarnemingen in de eerste steekproef met rangnummer $R_{Nn} > (N+1)/2$.

3.8. We merken op dat de toets van Fisher-Yates asymptotisch uniform meest onderscheidend is voor het toetsen van $\{H_{0N}\}$ tegen rijen normale verschuivingsalternatieven met bijbehorende scorefunctie $\gamma_0(t) = \Phi^{-1}(t)$; $0 < t < 1$, en dat de toets van Wilcoxon asymptotisch uniform meest onderscheidend is voor het toetsen van $\{H_{0N}\}$ tegen rijen logistieke verschuivingsalternatieven met bijbehorende scorefunctie $\gamma_1(t) = 2t - 1$. We passen stelling 3.18 toe. Nu geldt $\int_0^1 \gamma_0(t)dt = \int_0^1 \gamma_1(t)dt = 0$. aan de Noether-voorwaarde is voldaan en in het twee-steekproeven probleem geldt $\rho_1 = 1$. Bovendien

$$\rho_2 = \frac{\int_0^1\gamma_0(t)\gamma_1(t)dt}{(\int_0^1\gamma_0^2(t)dt\int_0^1\gamma_1^2(t)dt)^{\frac{1}{2}}} = \frac{\int_0^1\Phi^{-1}(t)(2t-1)dt}{(\int_0^1(\Phi^{-1}(t))^2dt\int_0^1(2t-1)^2dt)^{\frac{1}{2}}} = \sqrt{\frac{3}{\pi}},$$

zodat de gevraagde asymptotische relatieve doeltreffendheid gelijk is aan $\rho^2 = \rho_1^2\rho_2^2$, hetgeen in geval a) oplevert $\pi/3$ en in geval b) $3/\pi$.

3.9. Zoals in opgave 3.8. behoeven we slechts ρ_2 te berekenen, voor de scorefuncties $\gamma_1(t) = 2t - 1$, $0 < t < 1$ (optimaal tegen logistieke verschuivingslaternatieven) en $\gamma_2(t) = \mathrm{sgn}(2t-1)$, $0 < t < 1$ (optimaal tegen dubbel-exponentiële verschuivingsalternatieven).
Nu geldt

$$\rho_2 = \frac{\int_0^1\gamma_1(t)\gamma_2(t)dt}{(\int_0^1\gamma_1^2(t)dt\int_0^1\gamma_2^2(t)dt)^{\frac{1}{2}}} = \frac{\int_0^1(2t-1)\mathrm{sgn}(2t-1)dt}{(\int_0^1(2t-1)^2dt\int_0^1\mathrm{sgn}^2(2t-1)dt)^{\frac{1}{2}}} = \frac{1/2}{(1/3)^{\frac{1}{2}}}.$$

De ARE van de toets van Wilcoxon t.o.v. de mediaantoets in het tweesteekproevenprobleem, voor het toetsen van $\{H_{ON}\}$ tegen rijen dubbelexponentiële verschuivingsalternatieven welke voldoen aan (3.43) en (3.44) voor zekere $\eta > 0$, $0 < \alpha < 1$, is dus gelijk aan $\rho^2 = \rho_1^2\rho_2^2 = \frac{3}{4}$.

3.10. a) Volgens het gegeven geldt voor iedere deelrij $\tilde{N}(N)$ met $(\tilde{N}(N)/N)^{\frac{1}{2}} \to c > 0$, als $N \to \infty$, dat voor alle $\eta > 0$,

$$\lim_{\tilde{N}\to\infty} \int \psi_{i\tilde{N}} dP^{\tilde{N}}_{\tau_{\tilde{N}}(\eta)} = 1 - \Phi(\xi_{1-\alpha} - \eta\delta_i), \quad i = 1,2.$$

Kies nu $\eta' = \eta(\tilde{N}(N)/N)^{\frac{1}{2}}$, dan geldt $\tau_{\tilde{N}}(\eta') = \tau_N(\eta)$

$$\lim_{\tilde{N}\to\infty} \int \psi_{i\tilde{N}} dP^{\tilde{N}}_{\tau_N(\eta)} = 1 - \Phi(\xi_{1-\alpha} - c\eta\delta_i), \quad i = 1,2.$$

b) Als

$$\lim_{N\to\infty} \int \psi_{1N} dP^{N}_{\tau_N(\eta)} = 1 - \Phi(\xi_{1-\alpha} - \eta\delta_1),$$

kies dan een rij natuurlijke getallen $\{\tilde{N}(N)\}$ zodanig dat $\lim_{N\to\infty} \tilde{N}(N)/N = (\delta_1/\delta_2)^2$. Nu geldt volgens a) inderdaad dat

$$\lim_{N\to\infty} \int \psi_{2\tilde{N}(N)} dP^{\tilde{N}(N)}_{\tau_N(\eta)} = 1 - \Phi(\xi_{1-\alpha} - (\delta_1/\delta_2)\eta\delta_2 = 1 - \Phi(\xi_{1-\alpha} - \delta_1\eta) =$$

$$= \lim_{N\to\infty} \int \psi_{1N} dP^{N}_{\tau_N(\eta)}.$$

De laatste relatie in de opgave is een onmiddellijk gevolg van de definitie van asymptotische relatieve doeltreffendheid (zie hoofdstuk 2, pag. 54).

LITERATUUR

[1] ANDERSON, T.W., *An Introduction to Multivariate Statistical Analysis*, J. Wiley and Sons (1958).

[2] BAUER, H., *Wahrscheinlichkeitstheorie und Grundzuge der Masstheorie*, Walter de Gruyter & Co. (1968).

[3] BEHNEN, K., *Asymptotic optimality and ARE of certain rank-order tests under contiguity*, Ann. Math. Statist. 42, 325-329 (1971).

[4] BEHNEN, K., *A characterization of certain rank-order tests with bounds for the asymptotic relative efficiency*, Ann. Math. Statist. 43, 1839-1851 (1972).

[5] BERAN, R., *Linear rank statistics under alternatives indexed by a vector parameter*, Ann. Math. Statist. 41, 1896-1905 (1970).

[6] BILLINGSLEY, P., *Convergence of Probability Measures*, Wiley & Sons, Inc., New York (1968).

[7] CHUNG, K.L., *A course on probability theory*, Brace and World, New York, Harcourt, (1968).

[8] FABIUS, J. & W.R. VAN ZWET, *Grondbegrippen van de waarschijnlijkheidsrekening*, 1970, MCS 10, Mathematisch Centrum, Amsterdam (1970).

[9] GNEDENKO, B.V. & A.N. KOLMOGOROV, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley (1954).

[10] HÁJEK, J., *Asymptotically most powerful rank-order tests*, Ann. Math. Statist. 33, 1129-1147 (1962).

[11] HÁJEK, J., *A course in Nonparametric statistics*, Holden-Day, San Francisco (1969).

[12] HÁJEK, J. & Z. ŠIDÁK, *Theory of Rank Tests*, Academic Press, New York and London, (1967).

[13] HEYER, H., *Allgemeine Grenzwertsätze der Wahrscheinlichkeitsrechnung*, in Studies in Mathematical Statistics, Theory and Applications, edited by K. Sarkadi en I. Vincze, 57-74, Akedémiai Kiadó, Budapest, 1968.

[14] KAKUTANI, S., *An equivalence of infinite product measures*, Ann. of Math. 49, 214-224 (1948).

[15] KRAFT, C., *Some conditions for consistency and uniform consistency of statistical procedures*, Univ. California Publ. Statist. 2 (1955), 125-241.

[16] LE CAM, L., *Locally asymptotically normal families of distributions*, Univ. California Pub. Statist., 3, 37-98, University of California Press (1960).

[17] LE CAM, L., *Théorie asymptotique de la décision statistique*, Les Presses de L'Université de Montreál.

[18] LE CAM, L., *Convergence of estimates under dimensionality restrictions*, Ann. Math. Statist. 1, no. 1, 38-53 (1973).

[19] LE CAM, L., *On the assumptions used to prove asymptotic normality of maximum likelihood estimates*, Ann. Math. Statist. 41, 802-828 (1970).

[20] LEHMANN, E.L., *Testing Statistical Hypothesis*, New York (1959).

[21] LOÈVE, M., *Probability Theory*, 3rd Edition, Van Nostrand (1963).

[22] ROUSSAS, G., *Contiguity of probability measures: some applications in statistics*, Cambridge University Press (1972).

[23] RUDIN, W., *Real and Complex Analysis*, McGraw-Hill, (1966).

[24] WITTING, H. & G. NOLLE, *Angewandte Mathematische Statistik*, B.G. Teubner, Stuttgart (1970).

SYMBOLEN

De naast de symbolen vermelde cijfers verwijzen naar de pagina waarop het betreffende symbool is ingevoerd.

$\mathcal{B}$ - σ-algebra van Borelverzamelingen van $\mathbb{R}$: 4.

$\mathcal{B}^k$ - σ-algebra van Borelverzamelingen van $\mathbb{R}^k$ (k>1):4.

$C_B(S)$ - klasse van reëelwaardige continue begrensde functies op S:4.

$C_C(S)$ - klasse van reëelwaardige functies of S met compacte drager : 4.

$N(\mu,\sigma^2)$ - normale verdeling met verwachting μ en variantie $\sigma^2(\geq 0)$:5.

$N(\mu,\nu;\ \sigma^2,\tau^2,\rho)$ - 2-dimensionale normale verdeling met verwachtingen μ en ν, varianties σ^2 en τ^2 en correlatiecoefficient ρ:5.

E - verwachting : 33.

Var - variantie : 33.

$L(X)$ - (defectieve) kansverdeling van X : 5.

Bi(N,p) - binomiale verdeling met parameters N en p : 141.

$\xrightarrow[v]{}$ - vage convergentie : 5.

$\xrightarrow[z]{}$ - zwakke convergentie : 7, 10.

$\xrightarrow[D]{}$ - convergentie in verdeling : 10.

$\xrightarrow[P_N]{}$ - convergentie in waarschijnlijkheid (bij kansmaten$\{P_N\}$) : 11.

Φ - naburig : 2, 12.

$\triangleleft$, $\triangleright$ - eenzijdig naburig : 12.

$\| P-Q \|$ - variatie-afstand van P en Q : 14.

$\delta(P,Q)$ - Hellinger-afstand van P en Q : 15.

I_N - : 17.

J_N - : 17.

L_N - aannemelijkheidsquotient : 2, 17.

Λ_N - logarithme van een aannemelijkheidsquotient : 18.

P,P_N,P_n,Q,Q_N,Q_n - kansmaten : 1, 12, 58

$P_\zeta,P_\theta,P_\tau,P_{\zeta(n,N)},P_{\theta(n,N)},P_{\tau(n,N)}$ - kansmaten geïndiceerd met parameter : 32, 33.

p_N,q_N - dichtheden van P_N, Q_N t.o.v. σ-eindige maat : 13.

$g_{\theta(N,n)},g_{\tau(N,n)}$ - dichtheden van $P_{\theta(n,N)}$, $P_{\tau(n,N)}$ t.o.v. σ-eindige maat : 33.

ϕ_N^* - meest onderscheidende toets : 2.

ζ,τ,θ - parameters : 32.

T - parameterruimte : 32.

J_0 - Fisher-informatie : 45.

ARE - asymptotische relatieve doeltreffendheid : 54.

$X^{(1)} = (X^{(1)},\ldots,X^{(N)})$ - geordende steekproef bij $X_1,\ldots,X_n$: 59.

$R_n = r_n(X_1,\ldots,X_N)$ - rangnummer van X_n in de rij $X_1,\ldots,X_N$: 59.

$R = (R_1,\ldots,R_N)$ - vector van rangnummers : 59.

F_C - : klasse van alle continue verdelingsfuncties op $\mathbb{R}$: 63.

F_S - : klasse van alle verdelingsfuncties op $\mathbb{R}$, symmetrisch om 0 : 63.

H_0,H_{0N} - hypothese van isomorie : 63, 86.

H_1,H_{1N} - hypothese van symmetrie : 64, 109.

H_2,H_{2N} - hypothese van onafhankelijkheid : 64, 109.

$(\alpha_N;\ \theta(N),\tau(N))$ - (eindig) toetsingsprobleem $P_{\theta(N)}$ tegen $P_{\tau(N)}$ te toetsen met onbetrouwbaarheidsdrempel α_N : 33.

(α,H,K) - (eindig) toetsingsprobleem $P^N \epsilon H$ tegen $P^N \epsilon K$ te toetsen met onbetrouwbaarheidsdrempel α : 67.

α,α_N - (asymptotische) onbetrouwbaarheidsdrempel : 67

$(\alpha;\{\theta(N)\},\{\tau(N)\})$ - asymptotisch toetsingsprobleem om, voor iedere N, $P_{\theta(N)}$ tegen $P_{\tau(N)}$ te toetsen met asymptotische onbetrouwbaarheidsdrempel α : 33.

$(\alpha;\{H_N\},\{K_N\})$ - asymptotisch toetsingsprobleem om, voor iedere N, $P^N \epsilon H_N$ tegen $P^N \epsilon K_N$ te toetsen met asymptotische onbetrouwbaarheidsdrempel α : 85.

ζ,Δ - lokatie parameter : 75, 77.

σ - schaal parameter : 76

$K_{c,\Delta}$ - enkelvoudig regressie alternatief : 69.

K_c - regressie alternatief : 69.

$K_c(\epsilon)$ - : 70.

$c_1,\ldots,c_n,\ c_{N1},\ldots,c_{NN}$ - regressieconstanten : 69, 86.

γ - score functie : 71.

$a_N(n,\gamma)$, $n = 1,\ldots,N$ exacte scores behorende bij functie γ : 71.

γ_0 - : 71.

γ_0^+ - 81.

$\bar{c},\bar{d},\bar{c}_N,\bar{d}_N$ - : 86, 104, 105.

$\bar{a}$ - : 90.

INDEX

aannemelijkheidsquotient - 17.
logarithme van - 18.
alternatieven
verschuivings - 77, 78, 79, 80, 82, 96, 98, 109.
LEHMANN - 81, 84, 99, 110.
regressie - 69, 86, 91, 105.
asymptotisch(e)
- meest onderscheidende rij toetsen 34.
- onbetrouwbaarheid 33, 34, 85.
- onderscheidingsvermogen 34.
- relatieve doeltreffendheid 32, 54, 55.
- toetsingsprobleem 33, 85.
- uniform meest onderscheidende rij toetsen 86, 99.

consistente rij toetsen 34.
contiguous : zie *naburig*
convergentie:
- in waarschijnlijkheid 11.
- in verdeling 10.
- vage - 5, 6.
zwakke - van (defectieve) kansmaten 7, 8.
zwakke - van (defectieve) verdelingsfuncties 10.
CRAMÉR - WOLD : lemma van - 10.

defectieve kansmaat (verdelingsfunctie) 4.

eindig toetsingsprobleem 33, 67.

FISHER - informatie 45, 91, 98.
FISHER-YATES - toets van 78.

GAUSS - toets van 129.

HELLINGER - *afstand* 15.
hypothese:
- H_0 (isometrie) 63, 64, 72, 77, 95, 99, 103, 105.
- H_1 (symmetrie) 64, 81, 109.
- H_2 (onafhankelijkheid) 65, 66, 84, 110.

LECAM:
eerste lemma van - 22.
tweede lemma van - 40.
derde lemma van - 28.
LEHMANN - *alternatieven* 81, 84, 99, 110.
lokaal meest onderscheidende rang toets - 69.
lokatieparameter - 75.

naburig (contiguous) 12.
eenzijdig - 12.
natuurlijk onbetrouwbaarheidsdrempel, 75.
NEYMAN-PEARSON: lemma van - 2.
NOETHER: - voorwaarde 105.
normale scores 78.
nulhypothese 36.

onbetrouwbaarheid: asymptotische - . 34.
onbetrouwbaarheidsdrempel : asymptotische - 33, 85.
onderscheidingsvermogen: asymptotisch - 34.

parameter 32.
probleem:
twee steekproeven - 77, 95, 99, 103, 146.
symmetrie - 81, 109, 146.
regressie - 84, 110.
k-steekproeven (tegen trend) - 72, 105, 145.

rangnummer 59
rangtoets

voor H_0 (isometrie) 64.
voor H_1 (symmetrie) 64.
voor H_2 (onafhankelijkheid) 64.
regressie alternatief 69, 86, 91, 105.
regressie constanten 69, 104.
relatief compact 9.
relatieve doeltreffendheid: asymptotische - 32, 54, 55

Scores
exacte - 71.
benaderde - 71, 102.
scorefunctie - 71.
steekproef - 60.
aselecte - 60.
geordende - 60.

tight: zie *uniform beperkt*
toets(en):
asymptotische meest onderscheidende rij - 34.
asymptotische uniform meest onderscheidende rij - 86, 99.
consistente rij - 34.
meest onderscheidende - 34.
meest onderscheidende rang - 67.
- van FISHER-YATES 78, 98, 146.
- van GAUSS 129.
- twee steekproeven t - 131.
- van SPEARMAN 84, 110.
- van VAN DER WAERDEN, 79, 103.
- van WILCOXON 79, 98, 99, 144, 146.
- 146.
teken - 83, 131, 146.
toetsingsprobleem
eindig - 33, 67.
asymptotisch - 33, 85.

uniform beperkt (Engels: *tight*) 9.

vage convergentie 5.
variatie-afstand 14.
verdeling: convergentie in - 10.
binomiale - 141.
Cauchy - 130, 144.
dubbel exponentiele - 83, 141, 145, 146
homogene - 87.
logistieke - 79, 83.
normale - 25, 79, 98, 103, 144, 146.
Poisson - 24.
standaard normale - 144.
standaard logistieke - 79, 144.
verdelingsvrij: 67.
verschuivingsalternatieven 77, 146.

Waarschijnlijkheid: convergentie in - 11.
VAN DER WAERDEN, toets van - 79, 103.
WILCOXON: symmetrie toets van - 83, 144, 146.
twee-steekproeven toets van - 79, 98, 99.

Zwakke convergentie 10.

UITGAVEN IN DE SERIE MC SYLLABUS

Onderstaande uitgaven zijn verkrijgbaar bij het Mathematisch Centrum, 2e Boerhaavestraat 49 te Amsterdam-1005, tel. 020-947272.

MCS 1.1 F. GÖBEL & J. VAN DE LUNE, *Leergang Besliskunde, deel 1: Wiskundige basiskennis*, 1965. ISBN 90 6196 014 2.

MCS 1.2 J. HEMELRIJK & J. KRIENS, *Leergang Besliskunde, deel 2: Kansberekening*, 1965. ISBN 90 6196 015 0.

MCS 1.3 J. HEMELRIJK & J. KRIENS, *Leergang Besliskunde, deel 3: Statistiek*, 1966. ISBN 90 6196 016 9.

MCS 1.4 G. DE LEVE & W. MOLENAAR, *Leergang Besliskunde, deel 4: Markovketens, en wachttijden*, 1966. ISBN 90 6196 017 7.

MCS 1.5 J. KRIENS & G. DE LEVE, *Leergang Besliskunde, deel 5: Inleiding tot de mathematische besliskunde*, 1966. ISBN 90 6196 018 5.

MCS 1.6a B. DORHOUT & J. KRIENS, *Leergang Besliskunde, deel 6a: Wiskundige programmering 1*, 1968. ISBN 90 6196 032 0.

MCS 1.7a G. DE LEVE, *Leergang Besliskunde, deel 7a: Dynamische programmering 1*, 1968. ISBN 90 6196 033 9.

MCS 1.7b G. DE LEVE & H.C. TIJMS, *Leergang Besliskunde, deel 7b: Dynamische programmering 2*, 1970. ISBN 90 6196 055 X.

MCS 1.7c G. DE LEVE & H.C. TIJMS, *Leergang Besliskunde, deel 7c: Dynamische programmering 3*, 1971. ISBN 90 6196 066 5.

MCS 1.8 J. KRIENS, F. GÖBEL & W. MOLENAAR, *Leergang Besliskunde, deel 8: Minimaxmethode, netwerkplanning, simulatie*, 1968. ISBN 90 6196 034 7.

MCS 2.1 G.J.R. FÖRCH, P.J. VAN DER HOUWEN & R.P. VAN DE RIET, *Colloquium stabiliteit van differentieschema's, deel 1*, 1967. ISBN 90 6196 023 1.

MCS 2.2 L. DEKKER, T.J. DEKKER, P.J. VAN DER HOUWEN & M.N. SPIJKER, *Colloquium stabiliteit van differentieschema's, deel 2*, 1968. ISBN 90 6196 035 5.

MCS 3.1 H.A. LAUWERIER, *Randwaardeproblemen, deel 1*, 1967. ISBN 90 6196 024 X.

MCS 3.2 H.A. LAUWERIER, *Randwaardeproblemen, deel 2*, 1968. ISBN 90 6196 036 3.

MCS 3.3 H.A. LAUWERIER, *Randwaardeproblemen, deel 3*, 1968. ISBN 90 6196 043 6.

MCS 4 H.A. LAUWERIER, *Representaties van groepen*, 1968. ISBN 90 6196 037 1.

MCS 5 J.H. VAN LINT, J.J. SEIDEL & P.C. BAAYEN, *Colloquium discrete wiskunde*, 1968. ISBN 90 6196 044 4.

MCS 6 K.K. KOKSMA, *Cursus ALGOL 60*, 1969. ISBN 90 6196 045 2.

MCS 7.1 *Colloquium Moderne rekenmachines, deel 1*, 1969. ISBN 90 6196 046 0.

MCS 7.2 *Colloquium Moderne rekenmachines, deel 2*, 1969. ISBN 90 6196 047 9.

MCS 8 H. BAVINCK & J. GRASMAN, *Relaxatietrillingen*, 1969. ISBN 90 6196 056 8.

MCS 9.1 T.M.T. COOLEN, G.J.R. FÖRCH, E.M. DE JAGER & H.G.J. PIJLS, *Elliptische differentiaalvergelijkingen, deel 1*, 1970. ISBN 90 6196 048 7.

MCS 9.2 W.P. VAN DEN BRINK, T.M.T. COOLEN, B. DIJKHUIS, P.P.N. DE GROEN, P.J. VAN DER HOUWEN, E.M. DE JAGER, N.M. TEMME & R.J. DE VOGELAERE, *Colloquium Elliptische differentiaalvergelijkingen, deel 2*, 1970. ISBN 90 6196 049 5.

MCS 10 J. FABIUS & W.R. VAN ZWET, *Grondbegrippen van de waarschijnlijkheidsrekening*, 1970. ISBN 90 6196 057 6.

MCS 11 H. BART, M.A. KAASHOEK, H.G.J. PIJLS, W.J. DE SCHIPPER & J. DE VRIES, *Colloquium Halfalgebra's en positieve operatoren*, 1971. ISBN 90 6196 067 3.

MCS 12 T.J. DEKKER, *Numerieke algebra*, 1971. ISBN 90 6196 068 1.

MCS 13 F.E.J. KRUSEMAN ARETZ, *Programmeren voor rekenautomaten; De MC ALGOL 60 vertaler voor de EL X8*, 1971. ISBN 90 6196 069 x.

MCS 14 H. BAVINCK, W. GAUTSCHI & G.M. WILLEMS, *Colloquium Approximatietheorie*, 1971. ISBN 90 6196 070 3.

MCS 15.1 T.J. DEKKER, P.W. HEMKER & P.J. VAN DER HOUWEN, *Colloquium Stijve differentiaalvergelijkingen, deel 1*, 1972. ISBN 90 6196 078 9.

MCS 15.2 P.A. BEENTJES, K. DEKKER, H.C. HEMKER, S.P.N. VAN KAMPEN & G.M. WILLEMS, *Colloquium Stijve differentiaalvergelijkingen, deel 2*, 1973. ISBN 90 6196 079 7.

MCS 15.3 P.A. BEENTJES, K. DEKKER, P.W. HEMKER & M. VAN VELDHUIZEN, *Colloquium Stijve differentiaalvergelijkingen, deel 3*, 1975. ISBN 90 6196 118 1.

MCS 16.1 L. GEURTS, *Cursus Programmeren, deel 1: De elementen van het programmeren*, 1973. ISBN 90 6196 080 0.

MCS 16.2 L. GEURTS, *Cursus Programmeren, deel 2: De programmeertaal ALGOL 60*, 1973. ISBN 90 6196 087 8.

MCS 17.1 P.S. STOBBE, *Lineaire algebra, deel 1*, 1974. ISBN 90 6196 090 8.

MCS 17.2 P.S. STOBBE, *Lineaire algebra, deel 2*, 1974. ISBN 90 6196 091 6.

MCS 17.3 N.M. TEMME, *Lineaire algebra, deel 3*, 1976. ISBN 90 6196 123 8.

MCS 18 F. VAN DER BLIJ, H. FREUDENTHAL, J.J. DE IONGH, J.J. SEIDEL & A. VAN WIJNGAARDEN, *Een kwart eeuw wiskunde 1946-1971, Syllabus van de Vakantiecursus 1971*, 1974. ISBN 90 6196 092 4.

MCS 19 A. HORDIJK, R. POTHARST & J.TH. RUNNENBURG, *Optimaal stoppen van Markovketens*, 1974. ISBN 90 6196 093 2.

MCS 20 T.M.T. COOLEN, P.W. HEMKER, P.J. VAN DER HOUWEN & E. SLAGT, *ALGOL 60 procedures voor begin- en randwaardeproblemen*, 1976. ISBN 90 6196 094 0.

MCS 21 J.W. DE BAKKER (red.), *Colloquium Programmacorrectheid*, 1975. ISBN 90 6196 103 3.

MCS 22 R. HELMERS, F.H. RUYMGAART, M.C.A. VAN ZUYLEN & J. OOSTERHOOF, *Asymptotische methoden in de toetsingstheorie toepassingen van naburigheid*, 1976. ISBN 90 6196 104 1.

MCS 23.1 J.W. DE ROEVER (red.), *Colloquium Onderwerpen uit de biomathematica, deel 1*, 1976. ISBN 90 6196 105 X.

* MCS 23.2 J.W. DE ROEVER (red.), *Colloquium Onderwerpen uit de biomathematica, deel 2*, 1976. ISBN 90 6196 115 7.

* MCS 24.1 P.J. VAN DER HOUWEN, *Numerieke integratie van differentiaalvergelijkingen, deel 1: Eenstapsmethoden* , 1974. ISBN 90 6196 106 8.

MCS 25 *Colloquium Structuur van Programmeertalen*, 1976. ISBN 90 6196 116 5.

MCS 26.1 N.M. TEMME (red.), *Nonlinear Analysis, volume 1*, 1976. ISBN 90 6196 117 3.

MCS 26.2 N.M. TEMME (red.), *Nonlinear Analysis, volume 2*, 1976. ISBN 90 6196 121 1.

MCS 27 M. BAKKER, P.W. HEMKER, P.J. VAN DER HOUWEN, S.J. POLAK & M. VAN VELDHUIZEN, *Colloquium Discreteringmethoden*, 1976. ISBN 90 6196 124 6.

* MCS 28 O. DIEKMAN, N.M. TEMME (EDS.), *Nonlinear Diffusion Problems*, 1976. ISBN 90 6196 126 2.

MCS 29.1 J.C.P. BUS (red.), *Numerieke Programmatuur, deel 1A, deel 1B*, 1976. ISBN 90 6196 128 9.

* MCS 29.2 J.C.P. Bus (red.), *Numerieke Programmatuur, deel 2*, ISBN 90 6196 144 0.

MCS 31 J.H. VAN LINT (red.), *Inleiding in de Coderingstheorie*, 1976. ISBN 90 6196 136 X.

MCS 32 L. GEURTS (red.), *Colloquium Bedrijfssystemen*, 1976. ISBN 90 6196 137 8.

* MCS 33 P.J. VAN DER HOUWEN, *Differentieschema's voor de berekening van waterstanden in zeeën en rivieren*, ISBN 90 6196 138 6.

* MCS 34 J. HEMELRIJK, *Orienterende cursus mathematische statistiek*, IBSN 90 6196 139 4.

De met een * gemerkte uitgaven moeten nog verschijnen.