STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Rapport S 265 (C 13)

Leergang Besliskunde

Hoofdstuk VI

Toetsingstheorie en aanpassing van verdelingen

door

G. Klerk-Grobben

(Gedeeltelijk een bewerking van de hoofdstukken 13 en 15 van rapport S 200 (C 8) door Prof. Dr J. Hemelrijk)

April 1960

1. Grondbegrippen

Men kan de meeste elementaire toepassingen van de statistiek globaal in twee groepen verdelen: toepassingen van de schattingstheorie, die in het vorige hoofdstuk behandeld is, en van de toetsingstheorie, die nu ter sprake komt. Het doel van de schattingstheorie is: op grond van steekproeven van waarnemingen tot quantitatieve schattingen van onbekende parameters of onbekende kansen te komen. De toetsingstheorie heeft daarentegen tot doel op grond van steekproeven te onderzoeken of één of meer onbekende parameters of kansen bepaalde gegeven waarden bezitten of niet.

In symbolen: is θ een onbekende parameter van een kansverdeling, dan geeft de schattingstheorie een schatting $\underline{t}$ voor θ met een kansverdeling die min of meer om θ geconcentreerd ligt, terwijl de toetsingstheorie er op gericht is te onderzoeken of θ al dan niet gelijk aan een bepaalde gegeven waarde θ_o kan zijn.

Er bestaat een nauw verband tussen beide theorieën. Zo is het intuïtief reeds duidelijk dat men, op grond van een gegeven waarnemingsreeks, tot verwerping van de getoetste waarde θ_o zal dienen over te gaan indien de schatting $\underline{t}$ voor θ die uit deze zelfde waarnemingsreeks gevonden wordt sterk van θ_o afwijkt. Bijv. wil men onderzoeken of de gemiddelde levensduur θ van een bepaald soort lampen gelijk aan θ_o=2000 uur kan zijn en vindt men uit een steekproef van 100 lampen een gemiddelde levensduur van 700 uur (met een niet te grote spreiding hieromheen), dan is $\theta = \theta_o = 2000$ niet acceptabel. Dit verband tussen beide theorieën zal in het volgende duidelijk naar voren komen.

Bij de toepassingen leidt een goede formulering van het probleem en het doel van het onderzoek meestal tot een duidelijke keus tussen beide methoden. Vergelijkt men bijv. twee geneesmiddelen (een klassiek probleem), wat betreft hun effect bij de bestrijding van een bepaalde ziekte, dan zal men het beste van de twee middelen gaan gebruiken zodra men de onderstelling van gelijkwaardigheid van deze middelen verworpen heeft. Toepassing van de toetsingstheorie is hier dan de aangewezen weg. Wenst men daarentegen de premie voor een bepaalde verzekering te berekenen, dan heeft men een quantitatieve schatting van het risico nodig en is het niet genoeg (of zelfs niet van belang) om te weten dat dit

risico groter is dan een bepaald bedrag of groter dan het risico dat aan een andere situatie verbonden is.

Bij de toetsingstheorie wordt weer een aantal nieuwe begrippen gebruikt, die we in algemene bewoordingen zullen definiëren en daarnaast zullen toelichten aan de hand van het toetsen van een kans (binomiale toetsen). Terwille van de overzichtelijkheid zullen we niet de meest algemene formulering van de toetsingstheorie geven. Met name zullen we in de loop van ons betoog een aantal onderstellingen invoeren, die in een volkomen algemene opzet niet thuis horen, maar waar men in de praktijk toch vaak van uitgaat.

Het waarnemingsmateriaal

Het waarnemingsmateriaal, dat uit één of meer waarnemingsreeksen kan bestaan, zullen we aangeven met

(1.1) $$\underline{w}_1, \underline{w}_2, \ldots, \underline{w}_n.$$

Deze stochastische grootheden behoeven in het algemeen niet onafhankelijk te zijn; wij zullen echter, zoals men ook in de praktijk vrijwel altijd doet, wel onderstellen dat $\underline{w}_1,\ldots,\underline{w}_n$ onderling onafhankelijk zijn. Verder behoeven $\underline{w}_1,\ldots,\underline{w}_n$ niet alle dezelfde verdeling te bezitten en kunnen ze ook meer dimensionaal verdeeld zijn; elke $\underline{w}_i$ kan bijv. een paar waarnemingen $(\underline{x}_i,\underline{y}_i)$ voorstellen.

Toegelaten hypothesen en te toetsen hypothese

Als de kansverdelingen der $\underline{w}_i$ volledig bekend waren, zou er geen toetsingsprobleem bestaan. In de regel is er over deze kansverdelingen wel iets bekend, d.w.z. er worden bepaalde onderstellingen over gemaakt, die als gegeven worden beschouwd. Hieronder valt meestal de onderstelling van onafhankelijkheid van de waarnemingen die we boven reeds vermelden en verder ook een onderstelling omtrent de vorm van de verdelingen. Zo berusten veel toetsen op de veronderstelling van normaliteit van de verdelingen. Wij zullen het geval beschouwen dat de verdelingen der $\underline{w}_i$ nog van één onbekende parameter afhangen maar overigens volledig bekend zijn. Geven wij de onbekende parameter aan met het symbool θ, dan kunnen we de simultane verdelingsfunctie van $\underline{w}_1,\ldots,\underline{w}_n$ voorstellen door

(1.2) $\qquad F(w_1,\ldots,w_n|\,\theta),$

een functie die bekend is op de waarde van θ na. Onderstellen we dat de $\underline{w}_i$ onderling onafhankelijk zijn, dan kan (1.2) volgens de definitie 3.1 van hoofdstuk III als produkt geschreven worden:

$$(1.3) \qquad F(w_1,\ldots,w_n|\,\theta) = \prod_{i=1}^{n} F_i(w_i|\,\theta).$$

Door aan θ een bepaalde waarde te geven is F geheel vastgelegd en bekend. Een hypothese omtrent θ die F volledig bepaalt wordt een <u>enkelvoudige hypothese</u> genoemd. Welke waarden we voor θ mogelijk achten, hangt af van de aard van de parameter θ. Is θ een kans dan kan θ alleen waarden tussen 0 en 1 aannemen. Stelt θ de parameter λ uit de gamma-verdeling voor (pg. 7, hoofdstuk III), dan moet $\theta > 0$ zijn. Is θ het gemiddelde van een normale verdeling, dan zijn alle waarden $-\infty<\theta<+\infty$ mogelijk. Bij iedere mogelijke waarde van θ hoort dus een enkelvoudige hypothese en de verzameling van al deze hypothesen wordt de verzameling van <u>toegelaten hypothesen</u> genoemd.

Veelal gaat het er nu om te onderzoeken of een enkelvoudige hypothese van de vorm

$$(1.4) \qquad \theta = \theta_o \ ,$$

waarin dus θ_o een gegeven getal is, op grond van de waarnemingen (1.1) al of niet verworpen moet worden. Deze hypothese heet dan de <u>te toetsen hypothese</u> of ook wel de <u>getoetste hypothese</u> of <u>nul-hypothese</u>, vaak aangegeven door het symbool H_o. De te toetsen hypothese kan ook een <u>samengestelde hypothese zijn</u>, bijv.

$$(1.5) \qquad H_o : \theta \leq \theta_o \quad \text{of} \quad H_o : \theta \geq \theta_o .$$

Doordat bij een nulhypothese van deze vorm de waarde van θ niet precies is vastgelegd, is ook de verdeling $F(w_1,\ldots,w_n|\,\theta)$ niet volledig bepaald. We zullen later zien dat ook in deze situatie de toets gebaseerd is op de verdeling $F(w_1,\ldots,w_n|\,\theta_o)$, die wel bepaald is.

De uitkomst van het onderzoek is, dat de getoetste hypothese H_o al dan niet <u>verworpen</u> wordt. Bij verwerping van $H_o:\theta = \theta_o$ luidt de conclusie uiteraard: $\theta \neq \theta_o$. Bij verwerping van $H_o:\theta \leq \theta_o$

is de conclusie $\theta > \theta_o$. De hypothesen die aanvaard worden als H_o verworpen wordt, worden de alternatieve hypothesen genoemd. Bij niet-verwerpen van H_o wordt, zoals we later zullen zien niet tot aanvaarding van H_o besloten. Althans dit is, strikt genomen, niet verantwoord. Niet verwerpen betekent nl. alleen dat de waarnemingen geen voldoende aanwijzingen tegen H_o leveren, doch dat houdt nog niet in dat de aanwijzingen voor H_o voldoende sterk zijn om tot aanvaarding van H_o over te gaan. Dit punt zal later duidelijker worden.

Voorbeeld 1.1

Fraaie voorbeelden van het optreden van kansen met bekende waarden vindt men in de erfelijkheidsleer. Aan de proeven van MENDEL ligt een erfelijkheidstheoretisch model ten grondslag, dat bijv. leidt tot de uitkomst, dat bij het kruisen van bepaalde planten één van de twee verschijnselen A en B zich voordoet en dat de kans op A bij iedere proef een bekende waarde (bijv. $\frac{1}{4}$) bezit. De proeven van MENDEL dienden nu om te onderzoeken of de waarde van zo'n kans, en daarmee de ten grondslag liggende erfelijkheidstheorie, juist kon zijn. De verschijnselen A en B kunnen bijv. twee kleuren (van erwten, of bloemen) zijn of iets dergelijks. De n waarnemingen, verkregen bij n onafhankelijke proeven van deze aard, nemen in dit geval de "waarden" A of B aan; men kan dit in cijfers omzetten door bij iedere proef het aantal malen dat A optreedt als waarneming (w) te beschouwen; deze kan dus 1 of 0 zijn. De waarnemingsreeks (1.1) is dan dus een rij van n énen en nullen. Iedere waarneming heeft de alternatief verdeling (zie pg. ? hoofdstuk III) met kans $p = P[A]$, de kans op het verschijnsel $A(w_i=1)$. Nu zal als p het doel van het onderzoek is en de proeven onafhankelijk zijn, de volgorde van de énen en nullen in de waarnemingsreeks niet belangrijk zijn, maar wel het totaal malen dat A bij de n proeven opgetreden is. Noemen we dit aantal $\underline{x}$ dan is dus

$$(1.6) \qquad \underline{x} = \sum_{i=1}^{n} \underline{w}_i \ .$$

In hoofdstuk III, voorbeeld 2;7 (pg. 18) werd reeds opgemerkt dat $\underline{x}$ een binomiale verdeling met parameters n en p bezit, dus

(1.7) $$P[\underline{x}=x|p] = \binom{n}{x} p^x q^{n-x} (x=0,1,\ldots,n;\ q=1-p) .$$

In dit voorbeeld speelt p de rol van de onbekende parameter θ. Omdat dit een kans voorstelt zal voor p moeten gelden:

$$0 \leq p \leq 1 .$$

Dit bepaalt de verzameling van toegelaten hypothesen. De nulhypothese luidt bijv.

$$H_o : p = \frac{1}{4}$$

of algemeen: $p = p_o$ (waarin p_o een getal tussen 0 en 1 is).

Verwerping van H_o betekent de conclusie $p \neq \frac{1}{4}$. In praktische bewoordingen: de erfelijkheidstheorie, die tot de waarde $p = \frac{1}{4}$ leidde is niet juist en moet door een andere vervangen worden. De nieuwe hypothese zal dan aan hernieuwde proefnemingen getoetst moeten worden.

<u>De toetsingsgrootheid en zijn verdeling voor $\theta = \theta_o$</u>

De beoordeling van de te toetsen hypothese geschiedt met behulp van een <u>toetsingsgrootheid</u>, die uit de waarnemingen berekend wordt:

$$\underline{t} = t(\underline{w}_1,\ldots,\underline{w}_n).$$

Deze grootheid kan in principe meer dimensionaal zijn, maar is gewoonlijk ééndimensionaal. Meestal is $\underline{t}$ een schatting van de onbekende parameter θ, of - wat hiermee aequivalent is - een monotone functie van een schatting van θ.

Is $\theta=\theta_o$, dan is de verdelingsfunctie F van $(\underline{w}_1,\ldots,\underline{w}_n)$ volledig bekend en valt daaruit die van $\underline{t}$ af te leiden. Notatie: $F(t|\theta_o)$.

Zowel bij enkelvoudige $H_o(\theta=\theta_o)$ als bij samengestelde nulhypothesen ($\theta \leq \theta_o$ of $\theta \geq \theta_o$) is de uitvoering van de toets grotendeels op deze verdeling van $\underline{t}$ gebaseerd.

<u>Voorbeeld 1.2</u>

Vroeger zagen we dat, voor een geval als bij de Mendelproeven, als schatting voor p gevonden wordt

$$\underline{\hat{p}} = \frac{\underline{x}}{n} = \sum \frac{\underline{w}_i}{n} \quad .$$

Bij de toetsing van de hypothese $p = \frac{1}{4}$ wordt nu

(1.8) $\underline{t} = \underline{x} = n\,\underline{\hat{p}}$

als toetsingsgrootheid gebruikt. De verdeling van $\underline{t}$ bij $p = \frac{1}{4}$ is

(1.9) $P\left[\underline{x}=x \mid p=\frac{1}{4}\right] = \binom{n}{x}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{n-x} \qquad (x = 0,1,\ldots,n).$

Eén- en tweezijdige toetsing

Naarmate de gevonden waarde, t, voor de toetsingsgrootheid $\underline{t}$ verder in de staarten van de verdeling van $\underline{t}$ bij $\theta=\theta_o$ terechtkomt zal men oordelen dat het waarnemingsmateriaal minder in overeenstemming is met $\theta=\theta_o$. Is $\underline{t}$ een schatting van θ of een monotone functie hiervan, zoals in voorbeeld 1.2 en wil men toetsen $H_o:\theta=\theta_o$, dan zullen zowel waarden van $\underline{t}$ ver in de rechterstaart als waarden ver in de linkerstaart aanleiding zijn H_o te verwerpen. Dit wordt tweezijdige toetsing genoemd. Is de nulhypothese $\underline{\theta}\leq\theta_o$ dan zal een schatting voor θ die veel groter is dan θ_o tot verwerping van H_o leiden, een schatting die kleiner is dan θ_o echter niet, want waarden van θ kleiner dan θ_o zijn nu ook in H_o opgenomen. De toets heet dan rechtséénzijdig. Bij $H_o:\theta \geq \theta_o$ zullen alleen schattingen voor θ die veel kleiner zijn dan θ_o verwerping van H_o tot gevolg hebben: linkséénzijdige toetsing.

Voorbeeld 1.3

Bij de Mendel-proeven, waar het gaat om de toetsing van een speciale waarde van p zal men dus tweezijdig toetsen. Vaak hangt de keus tussen één of tweezijdige toetsing af van de consequenties die een verwerpen van H_o met zich meebrengt. Om dit te illustreren geven wij het volgende voorbeeld.

Een bedrijf overweegt de aanschaf van een nieuwe machine, maar men wil hiertoe alleen overgaan indien de nieuwe machine beter is dan de oude. Van de oude machine is bekend dat deze een fractie p_o aan uitval geeft. Om tot een beslissing te komen toetst men de hypothese $H_o:p \geq p_o$ (waarin p de fractie uitval van de nieuwe machine voorstelt). Alleen bij verwerping van H_o gaat men tot aanschaf van de nieuwe machine over. Een tweezijdige toets zou

hier niet op zijn plaats zijn, daar verwerping van $H_o: p=p_o$ ten gunste van $p > p_o$ evenmin tot aanschaf van de nieuwe machine leidt als niet-verwerpen van H_o.

Sommige toetsen, zoals de χ^2-toets die we in paragraaf 4 zullen bespreken, zijn essentieel éénzijdig. De χ^2-toets geeft bij iedere afwijking van H_o een vergrote kans op grote waarden van de toetsingsgrootheid, deze toets moet dan ook steeds rechts-éénzijdig gebruikt worden. De toetsingsgrootheid is hier geen schatting van een parameter.

De overschrijdingskans

Om te kunnen beslissen of men H_o zal verwerpen berekent men de overschrijdingskans bij de gevonden waarde t van $\underline{t}$, d.w.z.: de rechteroverschrijdingskans bij rechtseenzijdige toetsing:

$$k_r \stackrel{def}{=} P[\underline{t} \geq t | \theta_o] \quad (1.10)$$

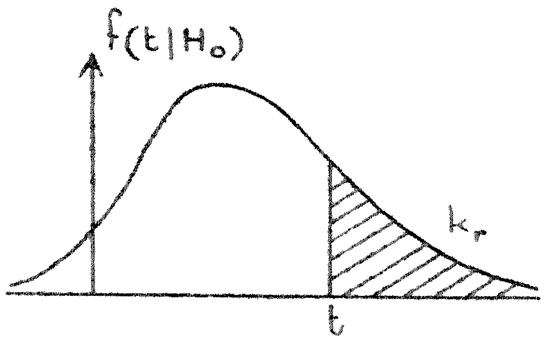
de linkeroverschrijdingskans bij linkseenzijdige toetsing:

$$k_\ell \stackrel{def}{=} P[\underline{t} \leq t | \theta_o] \quad (1.11)$$

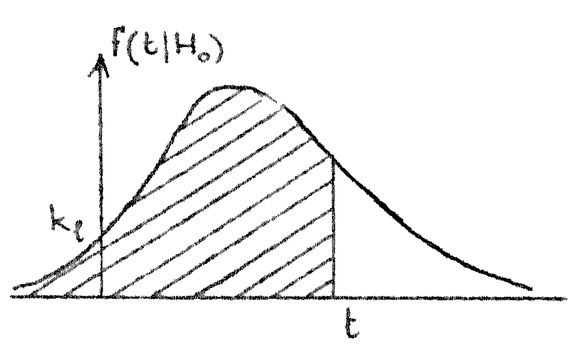
of de tweezijdige overschrijdingskans bij tweezijdige toetsing:

$$k \stackrel{def}{=} 2 \min (k_r, k_\ell) \quad (1.12)$$

In de meetkundige voorstelling van de verdelingsdichtheid van $\underline{t}$ onder H_o (zie figuur 1.1) komen deze kansen overeen met de gearceerde oppervlakken.



rechtseenzijdig



linkseenzijdig

fig. 1.1

Overschrijdingskansen bij gevonden waarde t van $\underline{t}$

De overschrijdingskansen zijn, elk bij de bijbehorende formulering van de nulhypothese, de kans om indien $\theta=\theta_o$ juist is, de gevonden waarde van t of nog een slechtere (d.w.z. nog minder in overeenstemming met H_o) te vinden. Bij de tweezijdige toets kunnen dit ook waarden uit de andere staart zijn dan die waarin t ligt, vandaar de factor 2 in de definitie van k. Heeft t een discrete verdeling, dan is de overschrijdingskans de som van de kansen van de t-waarden in de gearceerde gedeelten, de gevonden t meegerekend (hier is dus $k_r + k_\ell > 1$ als $P[\underline{t}=t|\theta_o] > 0$ is). Uit de definities (1.10) en (1.11) volgt $k_\ell = F(t|\theta_o)$ en

$$k_r = P[\underline{t} > t|\theta_o] + P[\underline{t}=t|\theta_o] = 1 - P[\underline{t} \leq t|\theta_o] + P[\underline{t}=t|\theta_o] =$$

$$= 1 - F(t|\theta_o) + P[\underline{t}=t|\theta_o].$$

Bij een continue verdeling van $\underline{t}$ is de laatste term gelijk aan 0.

Voorbeeld 1.4

Stel dat men bij Mendel-proeven 50 proeven doet om de hypothese $H_o: p=\frac{1}{4}$ te onderzoeken. De toetsingsgrootheid $\underline{x}$ (dit is dus het aantal malen dat A bij de 50 proeven is opgetreden) heeft dan onder H_o de verdeling (1.9) met n=50. $\underline{x}$ kan dus de waarden 0,1,.. ,50 aannemen. In tabel 1.I is de kans op elk van deze waarden opgenomen, voorzover ze niet te klein zijn. Verder zijn vermeld de linker overschrijdingskans $k_\ell = F(x)$ en de rechteroverschrijdingskans, die hier $k_r = 1-F(x-1)$ wordt.

Tabel 1.I

Kansen en overschrijdingskansen van alle waarden uit de binomiale verdeling met n=50 en $p=\frac{1}{4}$.

x	$P[\underline{x}=x]$	$k_l = F(x)$	$k_r = 1-F(x-1)$
0	0,000001	0,000001	1
1	0,000009	0,00001	≈ 1
2	0,00009	0,0001	≈ 1
3	0,0004	0,0005	0,9999
4	0,0015	0,002	0,9995
5	0,005	0,007	0,998
6	0,012	0,019	0,993
7	0,026	0,045	0,981
8	0,047	0,092	0,955
9	0,072	0,164	0,908
10	0,098	0,262	0,836
11	0,120	0,382	0,738
12	0,128	0,511	0,618
13	0,127	0,637	0,489
14	0,111	0,748	0,363
15	0,089	0,837	0,252
16	0,065	0,902	0,163
17	0,043	0,945	0,098
18	0,026	0,971	0,055
19	0,015	0,986	0,029
20	0,008	0,994	0,014
21	0,003	0,997	0,006
22	0,002	0,999	0,003
23	0,0006	0,9996	0,001
24	0,0002	0,9998	0,0004
25	0,0001	0,9999	0,0002
...,50	$\approx 0,0000$	≈ 1	$\approx 0,000$

(Deze tabel is ontleend aan H.G. ROMIG, 50-100 Binomial Tables, Wiley, N.Y., Chapman and Hall, London, 1953).

Is bij de 50 proeven 19 maal A opgetreden, dus is x=19, dan zen we in de tabel af dat $k_l = 0,986$ en $k_r = 0,029$, dus de tweejdige overschrijdingskans is $k = 2k_r = 0,058$.

Kritiek gebied en onbetrouwbaarheid

Bij een kleine overschrijdingskans zal men H_o verwerpen, bij een grote overschrijdingskans niet. Bij de uitvoering van de toets moet nu van tevoren vastgelegd worden, welke waarden van de overschrijdingskans nog aanvaardbaar zullen zijn en welke niet meer. Dat wil dus zeggen dat een grens, meestal aangeduid met α_o, gekozen wordt met de afspraak: H_o wordt verworpen als de overschrijdingskans bij $t \leq \alpha_o$ is en niet-verworpen indien de overschrijdingskans $> \alpha_o$ is. Vaak wordt gekozen $\alpha_o = 0,05$ of $\alpha_o = 0,01$. Is α_o gekozen, dan kan uit de verdeling van $\underline{t}$ onder θ_o bepaald worden welke waarden van $\underline{t}$ een overschrijdingskans $\leq \alpha_o$ zullen opleveren en welke dit niet doen. Zo wordt bij een rechtseenzijdige toets een <u>rechter-kritieke waarde</u> t_r gevonden, waarvoor geldt: t_r is de kleinste waarde van t waarvoor nog net

$$P[\underline{t} \geq t_r | \theta_o] \leq \alpha_o \qquad (1.13)$$

geldt. Heeft $\underline{t}$ een continue verdeling, dan kan t_r zo gekozen worden dat in (1.13) het =teken optreedt. Bij een discrete verdeling moet men ermee tevreden zijn

$$P[\underline{t} \geq t_r | \theta_o] = \alpha \text{ zo groot mogelijk} \leq \alpha_o \qquad (1.14)$$

te maken. Voor iedere waarde $> t_r$ wordt de overschrijdingskans dus zeker $< \alpha_o$. Op deze manier is een gebied, <u>het rechter kritieke gebied</u> Z_r geconstrueerd, bestaande dus uit alle waarden van $\underline{t}$ die tot verwerping van H_o zullen leiden.

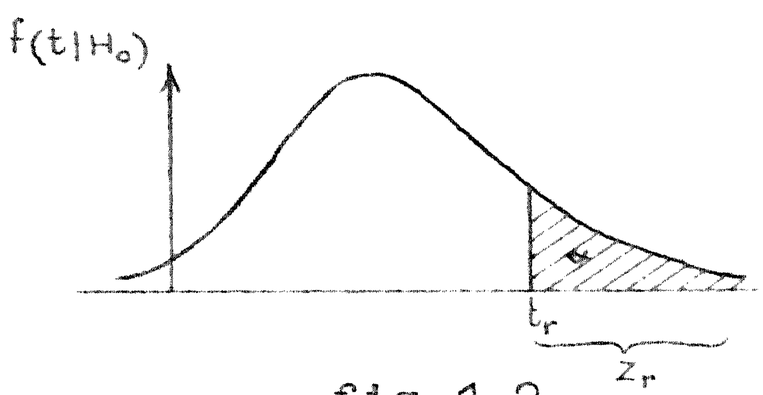


fig.1.2

<u>Rechter kritiek gebied</u>

Echter ook als $H_o(\theta \leq \theta_o)$ juist is zal $\underline{t}$ in het algemeen een positieve kans[1] bezitten om in het gebied Z_r terecht te komen. Bij $\theta = \theta_o$ is deze kans α (zie (1.14)). Is $\theta < \theta_o$ dan zal de kans dat $\underline{t}$ in Z_r komt $\leq \alpha$ zijn, omdat de werkelijke verdeling van $\underline{t}$, $F(t|\theta)$, in 't algemeen naar links verschoven zal zijn t.o.v. $F(t|\theta_o)$ (Z_r is

---

1) In hoofdstuk VII, paragraaf 2, zal een geval besproken worden waarbij deze kans =0 is.

is met behulp van $F(t|\theta_o)$ bepaald). Een waarde t in Z_r leidt tot verwerping van H_o en als H_o juist is, is dit een foute conclusie. Men noemt deze foute conclusie een fout van de eerste soort. De kans op een fout van de eerste soort is dus $\leq \alpha$ en $= \alpha$ als $\theta = \theta_o$. Deze kans wordt de onbetrouwbaarheid van de toets genoemd. We zagen al dat bij een continue verdeling van $\underline{t}$ geldt $\alpha = \alpha_o$ en bij een discrete in het algemeen $\alpha \leq \alpha_o$, als α_o de van tevoren vastgelegde grens voor de overschrijdingskansen is. α_o wordt de onbetrouwbaarheidsdrempel genoemd. De kans op het ten onrechte verwerpen van H_o kan verkleind worden door het gebied Z_r te verkleinen. Dit zal echter meestal tot gevolg hebben dat ook de kans op verwerping van H_o, als H_o inderdaad onjuist is, kleiner wordt. De keuze van Z_r en dus van α_o zal daarom op een compromis berusten. De onbetrouwbaarheid α(1.14) kan ook uitgedrukt worden als

$$(1.15) \qquad \alpha = P[\underline{t} \in Z_r | \theta_o] .$$

Bij linkseenzijdige toetsing wordt op dezelfde manier een linker kritieke waarde t_ℓ en een linker kritiek gebied Z_ℓ gevonden Is de verdeling van $\underline{t}$ bij $\theta = \theta_o$ symmetrisch om t=0 dan is $t_\ell = -t_r$.

Bij tweezijdige toetsing kunnen twee waarden t'_ℓ en t'_r bepaald worden zodat

$$(1.16) \quad P[\underline{t} \leq t'_\ell | \theta_o] \leq \tfrac{1}{2}\alpha_o \quad \text{en} \quad P[\underline{t} \geq t'_r | \theta_o] \leq \tfrac{1}{2}\alpha_o$$

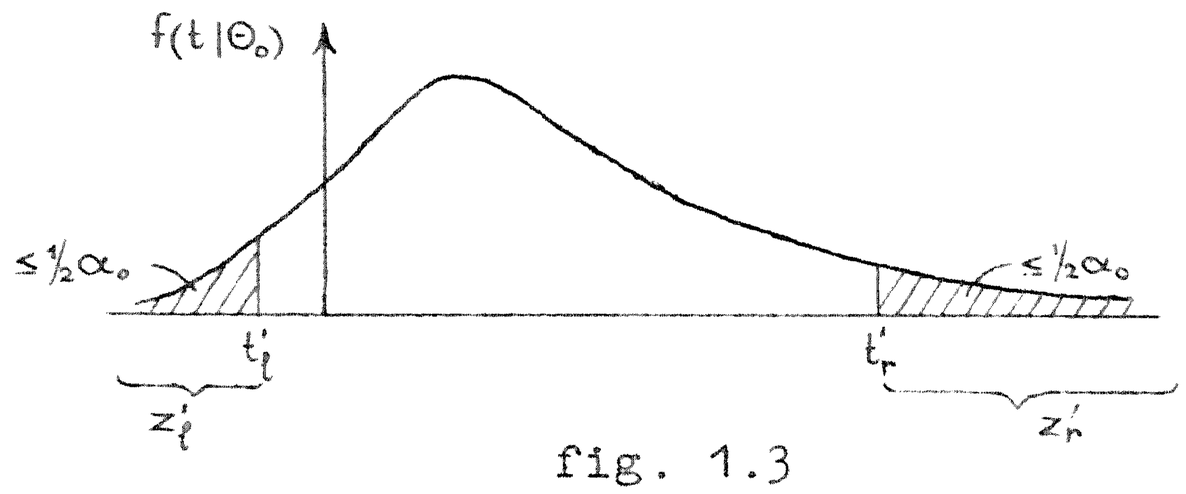


fig. 1.3

Tweezijdig kritiek gebied is $Z = Z'_\ell + Z'_r$

Op deze manier opgebouwd is dus een tweezijdige toets een combinatie van twee eenzijdige toetsen elk met onbetrouwbaarheidsdrempel $\frac{1}{2}\alpha_o$. Dit wordt een symmetrische tweezijdige toets genoemd. Het kritieke gebied is $Z = Z'_\ell + Z'_r$ (fig. 1.3). Men kan ook andere kritieke waarden kiezen door de eisen (1.16) te veranderen in

(1.17) $P[\underline{t} \leq t_\ell'' | \theta_o] \leq \alpha_o'$ en $P[\underline{t} \geq t_r'' \theta_o] \leq \alpha_o''$ met $\alpha_o' + \alpha_o'' = \alpha_o$,

de definitie van tweezijdige overschrijdingskans moet dan eveneens gewijzigd worden. Wij gaan hier niet verder op in.

De onbetrouwbaarheid van de tweezijdige toets is

(1.18) $$\alpha = P[\underline{t} \in Z | \theta_o] = P[\underline{t} \in Z_\ell' | \theta_o] + P[\underline{t} \in Z_r' | \theta_o],$$

dus de som van de onbetrouwbaarheden van de eenzijdige toetsen.

Voorbeeld 1.5

Kiezen we voor de toetsing van $p=\frac{1}{4}$ in een serie van n proeven als onbetrouwbaarheidsdrempel $\alpha_o=0,05$, dan vinden we met behulp van tabel 1.I als linkerdeel Z_ℓ' van het kritieke gebied Z de waarden $x = 0,1,\ldots,6$ en als rechterdeel Z_r' de waarden $x = 20,21,\ldots,50$. Een waarde voor $\underline{x}$ in één van deze gebieden zal dus aanleiding zijn $H_o: p=\frac{1}{4}$ te verwerpen. We kunnen dit nog iets nader preciseren tot: een waarde $x \in Z_\ell'$ zal aanleiding zijn $p=\frac{1}{4}$ te verwerpen ten gunste van de alternatieve hypothesen $p < \frac{1}{4}$ en een waarde $x \in Z_r'$ aanleiding om $p=\frac{1}{4}$ te verwerpen ten gunste van $p > \frac{1}{4}$. De onbetrouwbaarheid van de toets is

$$\alpha = P[\underline{x} \in Z_\ell' | p=\tfrac{1}{4}] + P[\underline{x} \in Z_r' | p=\tfrac{1}{4}] = 0,019 + 0,014 = 0,033 \ (< 0,05).$$

Onderscheidingsvermogen

Is in werkelijkheid H_o niet juist maar bijv. $\theta=\theta'$, dan kan de kans ons voor $\underline{t}$ een waarde in het kritieke gebied Z te vinden, aangegeven worden door

(1.19) $$\pi(\theta') \overset{\text{def}}{=} P[\underline{t} \in Z | \theta'].$$

Deze kans, dus de kans om H_o te verwerpen als θ' de juiste parameterwaarde is, wordt het onderscheidingsvermogen van de toets voor H_o t.o.v. $\theta=\theta'$ genoemd.

De kans om voor $\underline{t}$ een waarde buiten het kritieke gebied te vinden, dus H_o niet te verwerpen heet wel de kans op een fout van de tweede soort. Hiervoor geldt dus:

(1.20) $$P[\underline{t} \notin Z | \theta'] = 1 - P[\underline{t} \in Z | \theta'] = 1 - \pi(\theta').$$

Hoe groter het onderscheidingsvermogen, hoe kleiner de kans op een fout van de tweede soort. Het is duidelijk dat $\pi(\theta')$ vergroot kan worden door het kritieke gebied te kiezen, dus door de onbe-

trouwbaarheid α van de toets te vergroten. Bij de keuze van α_o, dus van α, zal men de belangrijkheid van de fouten van 1^e en 2^e soort tegen elkaar moeten afwegen. Een andere mogelijkheid om het onderscheidingsvermogen bij dezelfde onbetrouwbaarheid te vergroten is meestal: het aantal waarnemingen uit te breiden.

Praktisch altijd wordt de nulhypothese tegen een hele serie alternatieve hypothesen getoetst (bijv. bij $H_o:\theta=\theta_o$ zijn de alternatieven alle $\theta'\neq\theta_o$) en krijgen we te maken met het onderscheidingsvermogen $\pi(\theta')$ als functie van θ'. In zeer veel gevallen is deze functie niet bekend of ingewikkeld bijv. bij de χ^2-toetsen. Wij zullen op deze problemen verder niet ingaan.

Uitvoering van een toets

Bij vele standaardtoetsen bestaan tabellen van kritieke waarden t_ℓ en t_r bij gegeven onbetrouwbaarheidsdrempels (meestal $\alpha_o=0,05$; $\alpha_o=0,025$ en $\alpha_o=0,01$) en hoeft men dus de overschrijdingskans niet op te zoeken. In andere situaties zal men de overschrijdingskans bij de gevonden waarde van de toetsingsgrootheid $\underline{t}$ moeten berekenen of benaderen, meestal met behulp van een tabel van de verdelingsfunctie van $\underline{t}$ of een benadering hiervoor. Een voorbeeld hiervan geven wij in de volgende paragraaf.

Samenvatting

De mogelijke conclusies bij de verschillende toetsen (als $\underline{t}$ een monotone functie van een schatting voor θ is) kunnen we als volgt samenvatten.

Linkseenzijdige toetsing $H_o:\theta\geq\theta_o$

1) als $k_\ell\leq\alpha$ dus $t\leq t_\ell$ dan H_o verwerpen ten gunste van $\theta<\theta_o$

2) als $k_\ell>\alpha$ dus $t>t_\ell$ dan H_o niet verwerpen.

Rechtseenzijdige toetsing $H_o:\theta\leq\theta_o$

1) als $k_r\leq\alpha$ dus $t\geq t_r$ dan H_o verwerpen ten gunste van $\theta>\theta_o$

2) als $k_r>\alpha$ dus $t<t_r$ dan H_o niet verwerpen.

Tweezijdige toetsing $H_o: \theta = \theta_o$

1) als $k \leq \alpha$ dus $t \leq t'_\ell$ of $t \geq t'_r$ dan H_o verwerpen ten gunste van $\theta < \theta_o$ of $\theta > \theta_o$

als $k > \alpha$ dus $t'_\ell < t < t_r$ dan H_o niet verwerpen.

2. Binomiale toetsen met normale benadering

In hoofdstuk IV (stelling 3.3 en figuren 3.1) werd reeds gevonden, dat de binomiale verdeling met behulp van de aangepaste normale verdeling benaderd kan worden. Op deze benadering kan nog een verfijning aangebracht worden, die bekend staat onder de naam continuïteitscorrectie en betrekking heeft op het feit dat de continue normale verdeling gebruikt wordt om de discrete binomiale te benaderen.

Zijn n en p de parameters van de binomiale verdeling, dus is

$$P[\underline{x}=x \mid n,p] = \binom{n}{x} p^x q^{n-x},$$

met $\mathcal{E}\underline{x} = np$ en $\sigma^2(\underline{x}) = npq$, dan wordt de normale verdeling met dezelfde verwachting en variantie voor de benadering gebruikt. Een discrete kans $P[\underline{x}=x]$ moet nu benaderd worden door een oppervlak onder de normale verdelingsdichtheid en het ligt voor de hand hiervoor het oppervlak boven het interval $(x-\frac{1}{2}, x+\frac{1}{2})$ te nemen, dus symmetrisch om de waarde x gelegen (fig. 2.1). Alleen voor $P[\underline{x}=0]$ en $P[\underline{x}=n]$ neemt men de staart er volledig bij, dus dan neemt men het oppervlak boven het interval $(-\infty;\frac{1}{2})$ resp. $(n-\frac{1}{2};\infty)$, omdat anders de som der benaderende kansen < 1 zou zijn.

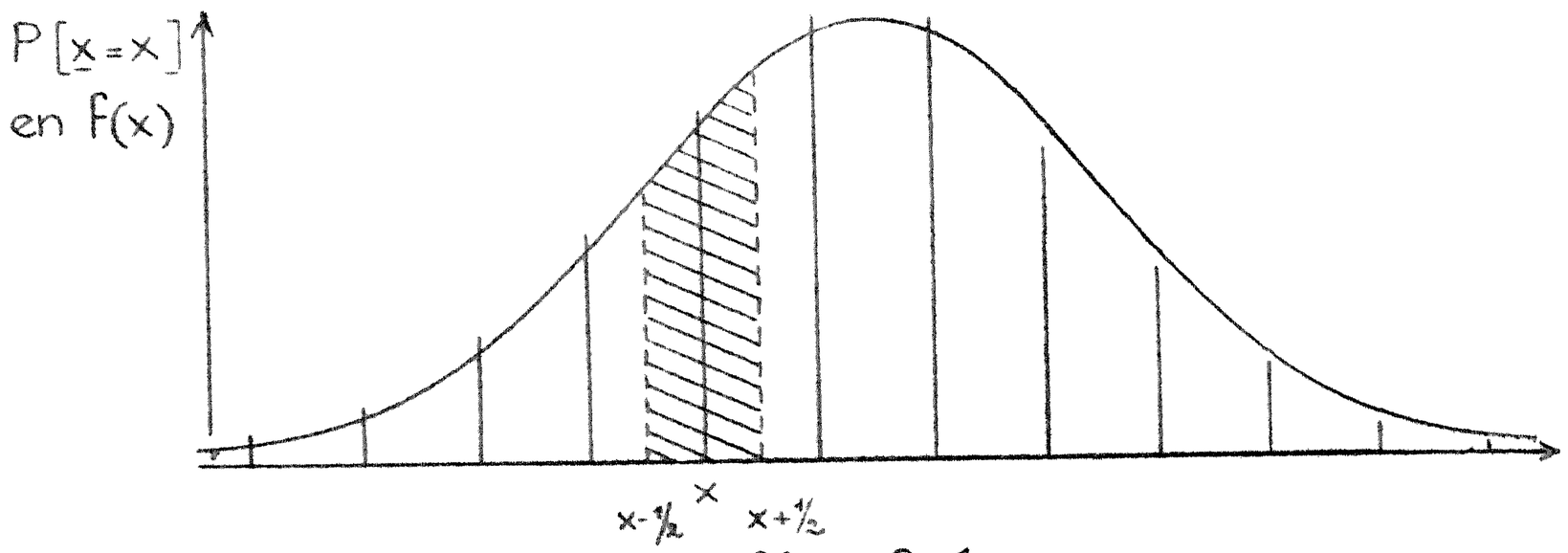


fig. 2.1

Normale benadering van een binomiale verdeling

Wenst men nu de <u>linker overschrijdingskans</u> $P\left[\underline{x} \leq x\right]$ te benaderen, dan neemt men daarvoor het "normale" oppervlak boven het interval $(-\infty,\ x+\frac{1}{2})$, terwijl voor de <u>rechter overschrijdingskans</u> $P\left[\underline{x} \geq x\right]$ dat boven $(x-\frac{1}{2};\infty)$ genomen wordt. Deze benaderende oppervlakken kunnen, na standaardisering (zie pg. 36, hoofdstuk III), uit tabel 4;II (hoofdstuk III) van de normale verdeling afgelezen worden. In formule wordt dit:

$$(2.1)\quad k_\ell(x) = P\left[\underline{x} \leq x \,\middle|\, n,p\right] \approx P\left[\underline{x} \leq x+\tfrac{1}{2} \,\middle|\, N(np;\ \sqrt{npq})\right] =$$

$$= P\left[\underline{u} \leq \frac{x-np+\frac{1}{2}}{\sqrt{npq}} \,\middle|\, N(0;1)\right],$$

$$(2.2)\quad k_r(x) = P\left[\underline{x} \geq x \,\middle|\, n,p\right] \approx P\left[\underline{x} \geq x-\tfrac{1}{2} \,\middle|\, N(np;\ \sqrt{npq})\right] =$$

$$= P\left[\underline{u} \geq \frac{x-np-\frac{1}{2}}{\sqrt{npq}} \,\middle|\, N(0;1)\right].$$

De tweezijdige overschrijdingskans is 2 min (k_r, k_ℓ), dus $2k_r$ als $x > np$ en $2k_\ell$ als $x < np$. Omdat $\underline{u}$ symmetrisch om 0 verdeeld is, is echter

$$P\left[\underline{u} \leq \frac{x-np+\frac{1}{2}}{\sqrt{npq}}\right] = P\left[\underline{u} \geq \frac{np-x-\frac{1}{2}}{\sqrt{npq}}\right],$$

zodat we voor k kunnen schrijven:

$$(2.3)\qquad k(x) \approx 2P\left[\underline{u} \geq \frac{|x-np|-\frac{1}{2}}{\sqrt{npq}}\right]$$

<u>Voorbeeld 2.1</u>

Voor $n=50$ en $p_o=\frac{1}{4}$ geldt volgens tabel 1.I bij $x=18$:

$$k_r = P\left[x \geq 18\right] = 0{,}055\ .$$

Met de normale benadering met continuïteitscorrectie wordt hiervoor gevonden ($\mathcal{E}\underline{x} = np_o = 12{,}5$; $\sigma(\underline{x}) = \sqrt{npq_o} = 3{,}06$)

$$k_r \approx P\left[\underline{x} \geq 18-0{,}5 \,\middle|\, N(12{,}5;\ 3{,}06)\right] = P\left[\underline{u} \geq \frac{18-12{,}5-0{,}5}{3{,}06}\right] =$$

$$= P\left[\underline{u} \geq 1{,}63\right] = 0{,}0516.$$

Zonder continuïteitscorrectie geeft de benadering:

$$k_r \approx P\left[\underline{u} \geq \frac{18-12,5}{3,06}\right] = P\left[\underline{u} \geq 1,79\right] = 0,0367.$$

Hier geeft dus de continuïteitscorrectie een betere benadering; dit is niet steeds het geval, maar wel geeft benadering met de correctie steeds een grotere uitkomst voor de overschrijdingskans dan zonder, zodat men H_o iets minder gauw zal verwerpen, dus iets "voorzichtiger" zal zijn. Bij grote n is de invloed van de correctie slechts gering.

Benadering voor kritieke waarden

Ook voor de kritieke waarden t_ℓ en t_r zijn nu eenvoudige benaderingsformules af te leiden. Is k_α de waarde in de N(0;1) verdeling waarvoor

(2.4) $$P\left[\underline{u} \geq u_\alpha\right] = \alpha$$

is (u_α is dus uit tabel 4;II, hoofdstuk III, te vinden bij $k_r = \alpha$), dan moet (bij benadering) voor de kritieke waarde x_r gelden

(2.5) $$u_\alpha = \frac{x_r - np_1 - \frac{1}{2}}{\sqrt{np_oq_o}} \quad \text{dus} \quad x_r \approx np_o - \tfrac{1}{2} + u_\alpha\sqrt{np_oq_o}$$

en voor x_ℓ

(2.6) $$-u_\alpha = \frac{x_\ell - np_o + \frac{1}{2}}{\sqrt{np_oq_o}} \quad \text{dus} \quad x_\ell \approx np - \tfrac{1}{2} - u_\alpha\sqrt{np_oq_o}\ .$$

Door combinatie hiervan met $u_{\frac{1}{2}\alpha}$ in plaats van u_α, vinden we bij de tweezijdige (symmetrische) toets de grenzen:

(2;7) $$x'_{r,\ell} \approx np_o \pm \left\{\tfrac{1}{2} + u_{\frac{1}{2}\alpha}\sqrt{np_oq_o}\right\}.$$

3. Betrouwbaarheidsgrenzen

De theorie der betrouwbaarheidsgrenzen vormt de verbinding tussen de schattingstheorie en de toetsingstheorie. Ook hier wordt het onderwerp van onderzoek gevormd door één (of meer) onbekende parameter(s) van een kansverdeling, waarover door een reeks (gewoonlijk onafhankelijke) waarnemingen $\underline{w}_1,\ldots,\underline{w}_n$ informatie wordt verschaft. Wij beperken ons tot de definitie en enkele voorbeelden om te laten zien hoe men van een toetsingsmethode zowel als een

schattingsmethode uitgaande, betrouwbaarheidsgrenzen kan construeren.

Definitie

Een functie $g_\ell = g_\ell(\underline{w}_1, \ldots, \underline{w}_n)$ van de waarnemingen is een betrouwbaarheids ondergrens voor een onbekende parameter θ, met onbetrouwbaarheidsdrempel α_ℓ, indien geldt

$$(3.1) \qquad P[\underline{g}_\ell < \theta] \geq 1 - \alpha_\ell ,$$

waarin θ de (onbekende) werkelijke waarde der parameter voorstelt.

Analoog geldt voor een betrouwbaarheids bovengrens met onbetrouwbaarheidsdrempel α_r:

$$(3.2) \qquad P[\underline{g}_r > \theta] \geq 1 - \alpha_r .$$

Tezamen vormen g_ℓ en g_r een betrouwbaarheidsinterval voor θ, met onbetrouwbaarheid α, indien geldt

$$(3.3) \qquad P[\underline{g}_\ell < \theta < \underline{g}_r] \geq 1 - \alpha;$$

hierin is $\alpha = \alpha_r + \alpha_\ell$ indien $P[\underline{g}_\ell < \underline{g}_r] = 1$.

Evenals bij toetsen kan onbetrouwbaarheidsdrempel vervangen worden door onbetrouwbaarheid indien de verdelingen van $\underline{g}_\ell$ en $\underline{g}_r$ continu zijn; de ongelijkheden in de definities gaan dan in gelijkheden over (natuurlijk niet binnen de vierkante haken).

Afleiding met behulp van de toetsingstheorie

Is er een toetsingsmethode T, waarmee op grond van de waarnemingen $\underline{w}_1, \ldots, \underline{w}_n$ iedere toegelaten (mogelijk geachte) waarde van de onbekende parameter getoetst kan worden, dan geldt de volgende stelling.

Is $\underline{t}$ de toetsingsgrootheid die bij de toetsingsmethode T gebruikt wordt, dan hangt de verdeling van $\underline{t}$ nog van de onbekende parameter θ af: $F(t|\theta)$. Denken we ons nu de toets T achtereenvolgens voor alle mogelijke waarden van θ, maar met hetzelfde waarnemingsmateriaal $w_1, \ldots, w_n$ en dezelfde onbetrouwbaarheidsdrempel α, uitgevoerd, dan zal de waarde t van $\underline{t}$ die alleen van $w_1, \ldots, w_n$ afhangt steeds dezelfde zijn, maar de verdeling van $\underline{t}$ en daarmee de overschrijdingskans van t met θ variëren. Bij deze serie toetsingen zullen nu sommige waarden van θ verworpen worden,

andere niet. Het gebied van mogelijke θ waarden wordt dus in twee delen gesplitst: $\underline{G}$ de verzameling van alle niet-verworpen θ-waarden en $\overline{\underline{G}}$ de verzameling van alle wel verworpen waarden. De splitsing hangt af van de waarnemingen: zouden we het hele proces met andere waarnemingen uitvoeren, dan zou in 't algemeen een andere splitsing verkregen worden. De splitsing heeft een kansverdeling, die afhangt van de verdeling van $\underline{w}_1,\ldots,\underline{w}_n$, vandaar de onderstreping van $\underline{G}$ en $\overline{\underline{G}}$. Voor $\underline{G}$ geldt nu

$$P[\theta \in \underline{G}] \geq 1-\alpha, \tag{3.4}$$

waarin θ de werkelijke parameterwaarde voorstelt; dit betekent dus dat $\underline{G}$ een betrouwbaarheidsgebied voor θ is. Het bewijs van (3.4) berust op de definitie van de onbetrouwbaarheidsdrempel bij een toets. Bij toetsing van een bepaalde waarde van θ heeft nl. deze θ-waarde een kans $\leq \alpha$ om verworpen te worden, indien deze θ-waarde de juiste is. Dit geldt dus ook voor de werkelijke waarde van θ die in de hele serie toetsingen ook een keer aan de beurt komt. Nu is verwerpen van een θ-waarde volgens bovenbeschreven procedure aequivalent met het weglaten van deze waarde uit $\underline{G}$. De werkelijke waarde van θ heeft dus ook een kans $\leq \alpha$ om uit $\underline{G}$ weggelaten te worden of anders gezegd: een kans $\geq 1-\alpha$ om wel in $\underline{G}$ opgenomen te worden, hetgeen in (3.4) wordt uitgedrukt.

De omgekeerde stelling geldt eveneens: Is $\underline{G}$ een betrouwbaarheidsgebied voor θ, dan kan hieruit een toets met dezelfde onbetrouwbaarheidsdrempel verkregen worden door de getoetste waarde θ_o te verwerpen indien θ_o niet in $\underline{G}$ ligt en niet te verwerpen indien θ_o wel in $\underline{G}$ ligt.

Voorbeeld 3.1

Daar met de binomiale toetsen iedere waarde van een onbekende kans p getoetst kan worden, kunnen hieruit ook betrouwbaarheidsgrenzen voor de onbekende kans geconstrueerd worden, eventueel door gebruik te maken van de benaderingsformules voor de overschrijdingskansen (vergelijk (2.1), (2.2) en (2.3). Het zou ons echter te ver voeren dit nader uit te werken. In plaats hiervan zullen we ter illustratie een eenvoudiger voorbeeld bekijken.

Stel dat de grootheid $\underline{x}$ een normale verdeling heeft met onbe-

kend gemiddelde μ en (bekende) spreiding 1. Om dit gemiddelde te onderzoeken worden n waarnemingen $\underline{x}_1,\ldots,\underline{x}_n$ verricht. Het gemiddelde, $\overline{\underline{x}} = \sum_{i=1}^{n} \frac{\underline{x}_i}{n}$, dat als toetsingsgrootheid gebruikt wordt, heeft de verdeling $N(\mu, \frac{1}{\sqrt{n}})$ (vgl. voorbeeld 3.1 uit hoofdstuk IV). Voor de parameter μ zijn nu alle waarden tussen $-\infty$ en $+\infty$ mogelijk en elke waarde kan getoetst worden. Bijv. $H_o: \mu = \mu_o$, met μ_o een gegeven getal, geeft als overschrijdingskansen bij de gevonden $\overline{x}$:

rechtseénzijdig

$$(3.5) \qquad k_r = P\left[\underline{\overline{x}} \geq \overline{x} \,|\, N(\mu_o, \tfrac{1}{\sqrt{n}})\right] = P\left[\underline{u} \geq (\overline{x} - \mu_o)\sqrt{n}\right],$$

linkseenzijdig

$$(3.6) \qquad k_l = P\left[\underline{\overline{x}} \leq \overline{x} \,|\, N(\mu_o, \tfrac{1}{\sqrt{n}})\right] = P\left[\underline{u} \leq (\overline{x} - \mu_o)\sqrt{n}\right]$$

en tweezijdig

$$(3.7) \qquad k = 2P\left[\underline{u} \geq |\overline{x} - \mu_o|\sqrt{n}\right].$$

Deze kansen kunnen in een tabel van de N(0,1) verdeling van $\underline{u}$ opgezocht worden. Gaan we na welke waarden van μ volgens deze toetsingsmethode, met onbetrouwbaarheidαen rechtseenzijdige toetsing, verworpen zouden worden, dan zien we uit (3.5) dat dit die waarden μ zijn die voldoen aan

$$(3.8) \qquad (\overline{x} - \mu)\sqrt{n} \geq u_\alpha,$$

waarin voor u_α geldt

$$P\left[\underline{u} \geq u_\alpha\right] = \alpha.$$

Immers dan is $k_r \leq \alpha$. Formule (3.8) kunnen we omvormen tot:

$$(3.9) \qquad \mu \leq \overline{x} - u_\alpha/\sqrt{n}.$$

Dit betekent dus dat

$$(3.10) \qquad \underline{g}_l = \underline{\overline{x}} - u_\alpha/\sqrt{n}$$

een betrouwbaarheids <u>ondergrens</u> voor μ is (met onbetrouwbaarheid α). Een <u>rechts</u>eenzijdige toetsing leidt dus tot een <u>ondergrens</u> voor μ.

Analoog vinden we bij linkseenzijdige toetsing een betrouwbaarheids <u>bovengrens</u>:

(3.11) $$\underline{g}_r = \overline{\underline{x}} + u_\alpha/\sqrt{n} \ .$$

Combinatie van deze twee geeft als betrouwbaarheidsinterval voor μ, met onbetrouwbaarheid 2α(!):

(3.12) $$\overline{\underline{x}} - u_\alpha/\sqrt{n} < \mu < \overline{\underline{x}} + u_\alpha/\sqrt{n} \ .$$

Afleiding van betrouwbaarheidsgrenzen met behulp van de schattingstheorie

Een andere methode om betrouwbaarheidsintervallen af te leiden berust op de schattingstheorie. Het principe van deze methode kan als volgt uiteengezet worden.

Laat $\underline{t}$ een schatting van de onbekende parameter θ zijn met $\mathcal{E}\underline{t} = \theta$ en $\sigma^2(\underline{t}) = \sigma^2$ onafhankelijk van θ. Laat verder de verdeling van $\underline{t}$, op de onbekende θ na, bekend zijn, bijv. normaal. Dan geldt

(3.13) $$P\left[\theta - u_{\frac{1}{2}\alpha} . \sigma < \underline{t} < \theta + u_{\frac{1}{2}\alpha} . \sigma\right] = 1 - \alpha .$$

Het interval tussen [ ] heet een voorspellingsinterval voor t; de voorspelling dat $\underline{t}$ in dit interval zal liggen heeft een kans $1-\alpha$ waar te zijn. Dit interval kan nu eenvoudig omgevormd worden tot een betrouwbaarheidsinterval voor θ. Immers (3.13) is aequivalent met

(3.14) $$P\left[\underline{t} - u_{\frac{1}{2}\alpha} . \sigma < \theta < \underline{t} - u_{\frac{1}{2}\alpha} \sigma\right] = 1 - \alpha .$$

Voorbeeld 3.2

In het vorige voorbeeld is $\overline{\underline{x}}$ een schatting voor μ met $\mathcal{E}\overline{\underline{x}} = \mu$ en $\sigma^2(\overline{\underline{x}}) = \frac{1}{n}$ onafhankelijk van μ. Toepassing van (3.14) geeft nu onmiddellijk een betrouwbaarheidsinterval voor μ van de vorm (3.12) (nu met onbetrouwbaarheid α omdat $u_{\frac{1}{2}\alpha}$ in plaats van u_α gebruikt is).

De eis, $\mathcal{E}\underline{t}=\theta$ en $\sigma^2(\underline{t})=\sigma^2$ onafhankelijk van θ, is niet noodzakelijk, maar werd gesteld om (3.13) eenvoudig in (3.14) te kunnen omzetten. Het kriterium voor de toepasbaarheid van deze methode ligt in de mogelijkheid om een voorspellingsinterval van de vorm (3.13) om te zetten in een betrouwbaarheidsinterval voor de onbekende parameter, hetgeen lang niet altijd mogelijk is.

Ook uit een interval op deze wijze geconstrueerd, kan weer een toets voor θ afgeleid worden.

4. Aanpassing van verdelingen

Dikwijls komt men bij statistischeanalyses te staan voor het probleem, dat men van een grootheid een aantal waarnemingen tot zijn beschikking heeft, maar om verder te kunnen werken, moet beslissen welke vorm de kansverdeling (eventueel met nog enkele onbekende parameters) van deze grootheid zal bezitten.

In sommige gevallen kan men op theoretische gronden tot een bepaalde vorm komen. Vaker echter zal men deze kansverdeling uit de waarnemingen moeten afleiden. Dit wordt dan het aanpassen van een verdeling genoemd. Gewoonlijk is dit een kwestie van proberen, waarvoor we enkele richtlijnen zullen aangeven, die echter niet volledig zijn.

Suggereren de waarnemingen een symmetrische verdeling, dan zal men een normale verdeling proberen aan te passen. Het gebruik van waarschijnlijkheidspapier (waarop we nog terugkomen) verschaft hierbij een vlugge grafische methode om de normaliteit van de waarnemingen te beoordelen.

Bij scheve verdelingen kan men onderzoeken of wellicht log $\underline{x}$ (als $\underline{x}$ de grootheid is die onderzocht wordt) een normale verdeling heeft, waarbij logarithmisch waarschijnlijkheidspapier gebruikt kan worden. Andere mogelijkheden bij scheve verdelingen zijn een exponentiële verdeling van $\underline{x}$ of van log $\underline{x}$ of een gammaverdeling.

Kan men met de grafische methoden niet tot een keus tussen verschillende mogelijke verdelingen komen dan bestaat er nog een andere methode nl. met behulp van χ^2-toetsen.

Beide methoden zullen nader besproken worden.

Waarschijnlijkheidspapier

We beschouwen eerst een stochastische grootheid $\underline{u}$ met de gestandaardiseerde normale verdeling (zie pg 6, hoofdstuk III) en geven de verdelingsfunctie aan met $\phi(u)$, dus

$$(4.1) \qquad \phi(u) = P[\underline{u} \leq u] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-v^2/2}\, dv.$$

De functie $\phi(u)$ op gewoon millimeterpapier tegen u uitgezet heeft een gedaante als in fig. 4.1 geschetst is.

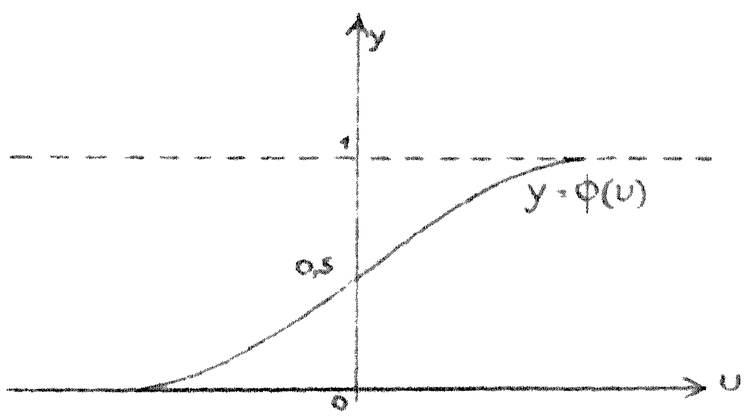


fig. 4.1

Gestandaardiseerde normale verdelingsfunctie

Door transformatie van de y-schaal in een z-schaal met $y=\Phi(z)$, gaat deze kromme over in de rechte lijn $z=u$, zie fig. 4.2. Bij $y=0,5$ behoort $z=0$, $y=1$ gaat over in $z=\infty$ en $y=0$ in $z=-\infty$, de staarten van de verdelingsfunctie komen nu dus niet meer op het papier. Op het waarschijnlijkheidspapier, dat in de handel verkrijgbaar is, is deze transformatie reeds uitgevoerd. Langs de z-schaal zijn hierop de bijbehorende waarden $y=\Phi(z)$ bijgeschreven. De y-waarden lopen gewoonlijk van 0,0002 tot 0,9998.

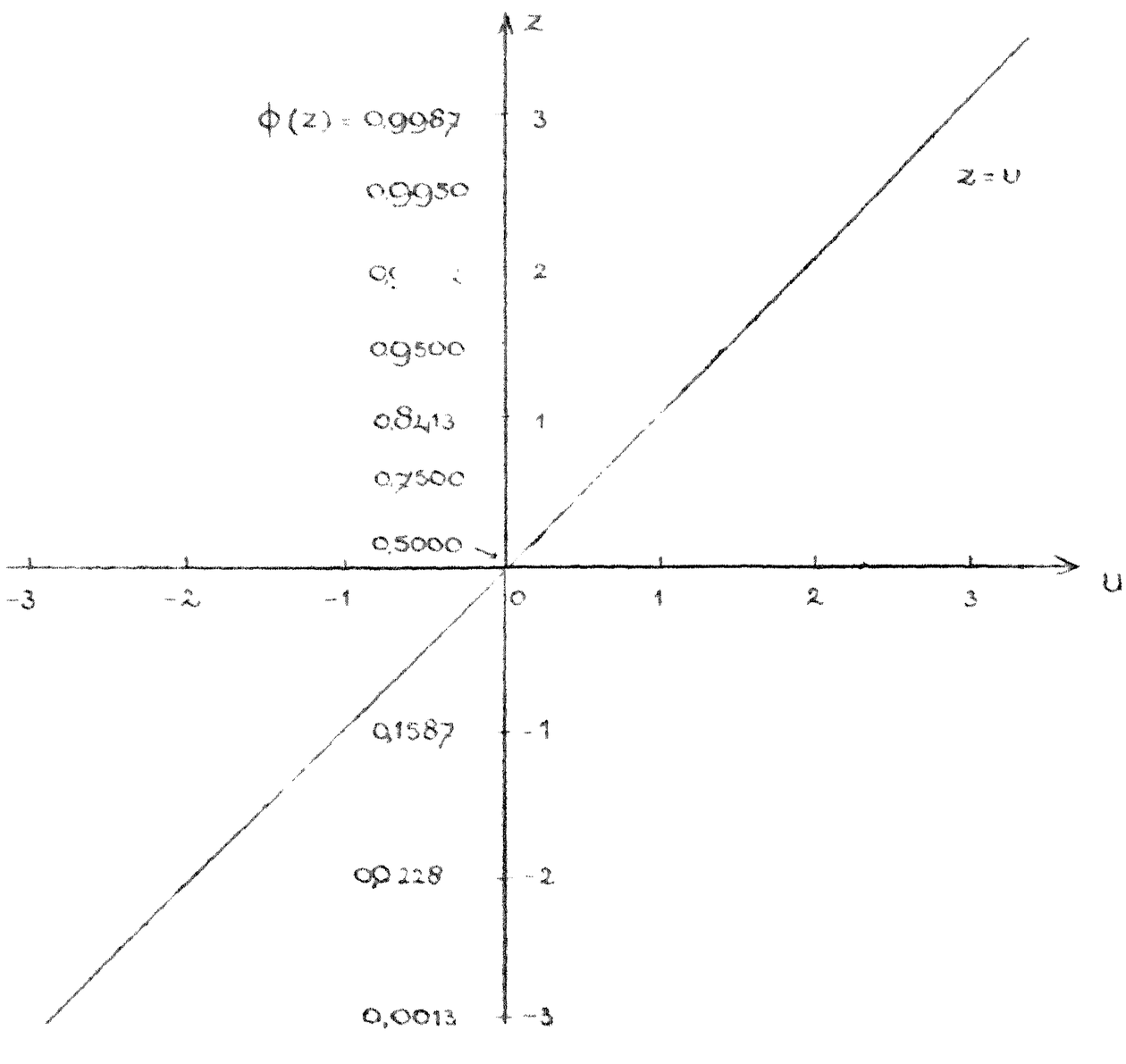


fig. 4.2

De gestandaardiseerde normale verdelingsfunctie op waarschijnlijkheidspapier

Bekijken we nu $\underline{x}$ met een normale verdeling met gemiddelde μ en spreiding σ, dan is de verdelingsfunctie:

$$(4.2) \quad f(x) = P\left[\underline{x} \leq x\right] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(v-\mu)^2/2\sigma^2}\, dv =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-w^2/2}\, dw = \phi\left(\frac{x-\mu}{\sigma}\right) ,$$

waarin $\phi(u)$ weer de verdelingsfunctie van de $N(0;1)$ verdeelde $\underline{u}$ voorstelt. $F(u)$ tegen x uitgezet heeft een vorm als in fig. 4.1, met dit verschil dat nu $y=0,5$ bij $x=\mu$ optreedt. Toepassing van de transformatie $y=\phi(2)$ maakt weer deze kromme tot een rechte lijn: $z=(x-\mu)/\sigma$, waarvan dus de plaats en de helling bepaald zijn door μ en σ (fig. 4.3): voor $x=\mu$ wordt $F(x) = \phi\left(\frac{x-\mu}{\sigma}\right)$ gelijk aan $\phi(0)=0,5$ en voor $x=\mu+\sigma$ wordt dit $\phi(1)=0,8413$.

Nu zijn echter μ en σ en dus ook $\phi\left(\frac{x-\mu}{\sigma}\right)$ onbekend; $\phi\left(\frac{x-\mu}{\sigma}\right)$ zal dan geschat moeten worden uit de waarnemingen $x_1, \ldots, x_n$ van $\underline{x}$. Om dit te doen worden de waarnemingen naar opklimmende grootte gerangschikt.

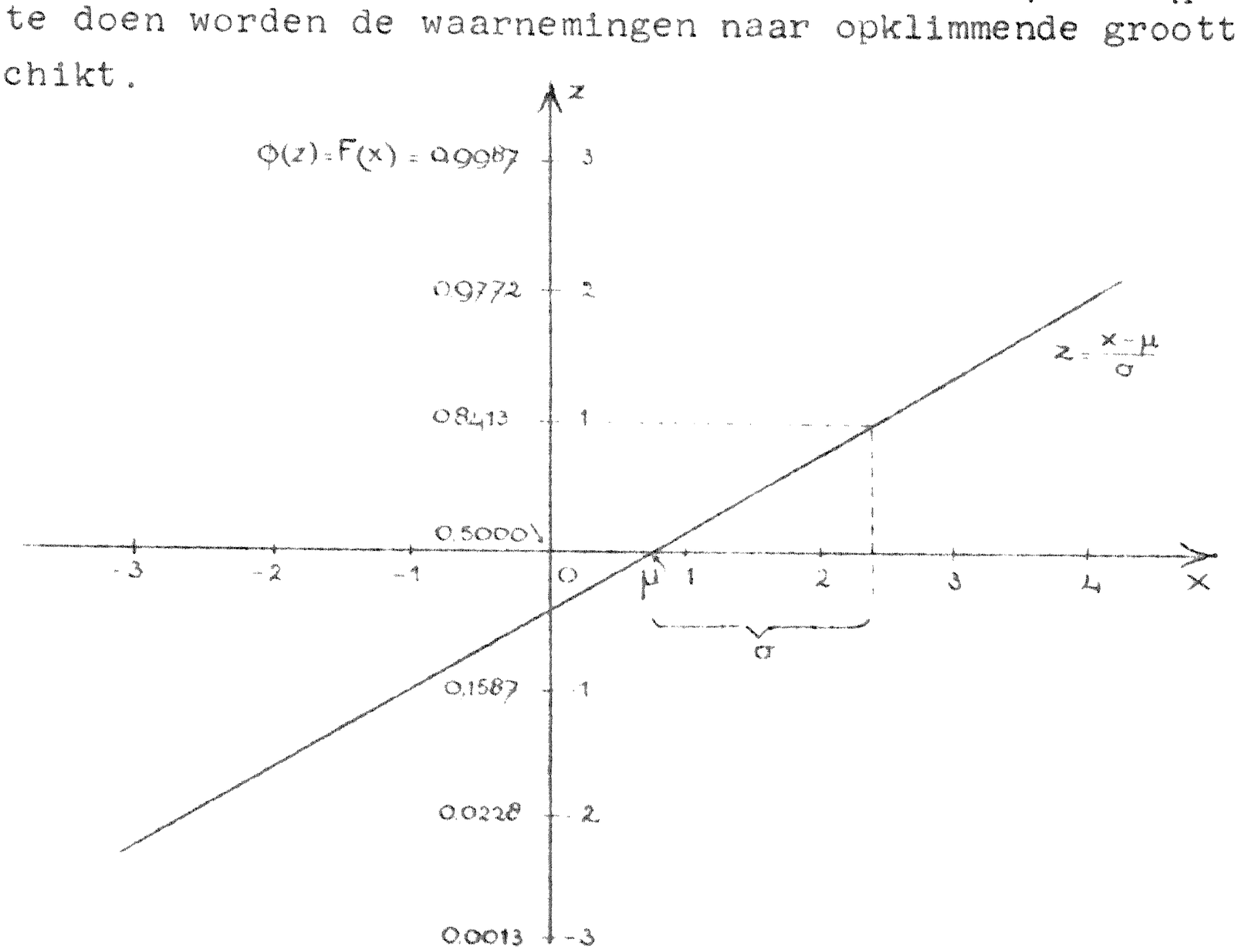


fig. 4.3

De normale verdelingsfunctie F(x) op waarschijnlijkheidspapier

Stel dat $x_1,\ldots,x_n$ reeds in deze volgorde staan. Een schatting voor $F(x_i)$ zou dan kunnen zijn $\frac{i}{n}$. Deze schatting heeft het bezwaar dat $F(x_n)$ door $\frac{n}{n}=1$ geschat wordt, d.w.z.: $z=\infty$ (want $\Phi(\infty)=1$), en dus buiten het papier valt. Dit is in vergelijking met de rechte lijn die de echte $F(x)$ oplevert, waar $z=\infty$ pas bij $x=\infty$ optreedt, een slechte schatting. De schatting $\frac{i}{n}$ kan op verschillende manieren iets veranderd worden om dit bezwaar te ondervangen. A. BENARD en E. BOS-LEVENBACH[1] vonden dat de beste eenvoudige schatting voor $y=F(x_i)$ gegeven wordt door

$$\hat{y} = \frac{i-0,3}{n+0,4} \quad . \tag{4.3}$$

Bij grote steekproeven kan men $\frac{i}{n}$ blijven gebruiken omdat dan de paar punten met kleine i en met $i \approx n$ geen belangrijke bijdrage tot het algemene beeld meer geven.

Zijn dus $x_1,\ldots,x_n$ de naar opklimmende grootte gerangschikte waarnemingen en liggen de punten $(x_i, \frac{i-0,3}{n+0,4})$ op het waarschijnlijkheidspapier ongeveer op een rechte lijn ("ongeveer" want $\frac{i-0,3}{n+0,4}$ is slechts een schatting voor $F(x_i)$), dan mag verondersteld worden dat $\underline{x}_1,\ldots,\underline{x}_n$ uit een normale verdeling afkomstig zijn.

Uit de zo goed mogelijk aangepaste rechte lijn kunnen schattingen voor μ en σ afgeleid worden. Voor μ is dit de x-waarde die volgens de rechte lijn bij $\Phi(z)=0,5$ behoort; σ volgt uit de helling van de lijn bijv. als verschil tussen de x-waarden bij $\Phi(z)=0,84$ en $\Phi(z)=0,5$. Deze schattingen zullen minder nauwkeurig zijn dan de gebruikelijke (vgl. (1.4) en (1.7) uit hoofdstuk V):

$$\underline{\bar{x}} = \Sigma \underline{x}_i/n \text{ en } \underline{s} = \sqrt{\Sigma(\underline{x}_i-\underline{\bar{x}})^2/(n-1)} \quad . \tag{4.4}$$

Systematische afwijkingen van een rechte lijn zullen er op wijzen dat de verdeling van $\underline{x}$ niet normaal is.

Zo ontstaat bij een verdeling waarvan de staarten dunner zijn dan bij de aangepaste normale verdeling op het waarschijnlijkheidspapier een kromme zoals in fig. 4.4 geschetst is. Zijn de staarten juist te dik dan ontstaat fig. 4.5. Fig. 4.6 en 4.7 schetsen het

---

1) A. BENARD en E.C. BOS-LEVENBACH, Het uitzetten van waarnemingen op waarschijnlijkheidspapier, Statistica 7(1953), 163-173.

resultaat bij scheve verdelingen resp. voor een verdeling scheef naar rechts en voor een verdeling scheef naar links; zie de frequentiedichtheden die erbij getekend zijn.

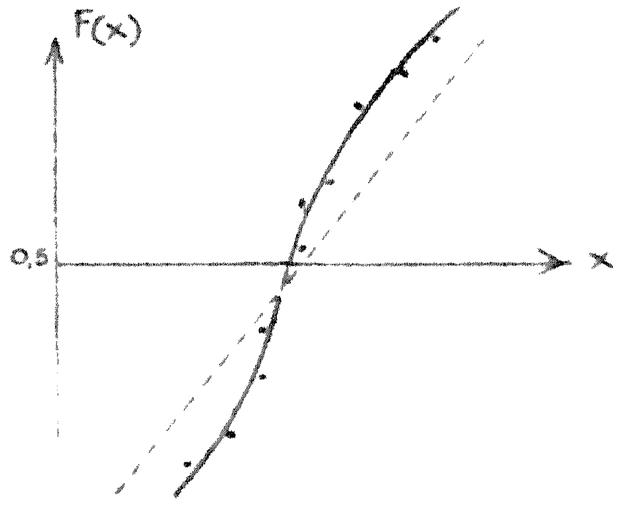


fig. 4.4
Staarten te dun

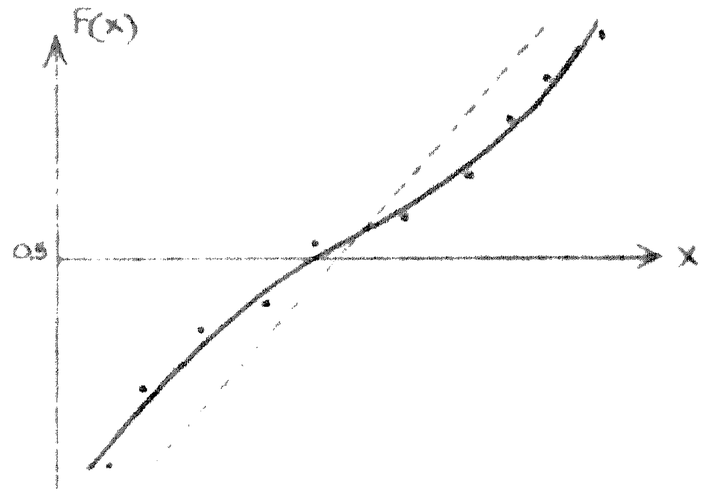


fig 4.5
Staarten te dik

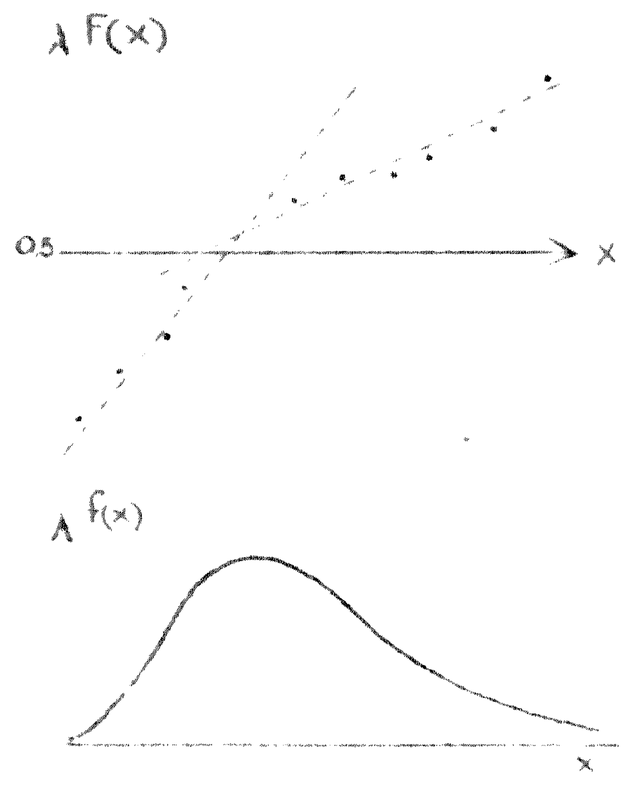


fig. 4.6
Verdeling scheef naar rechts

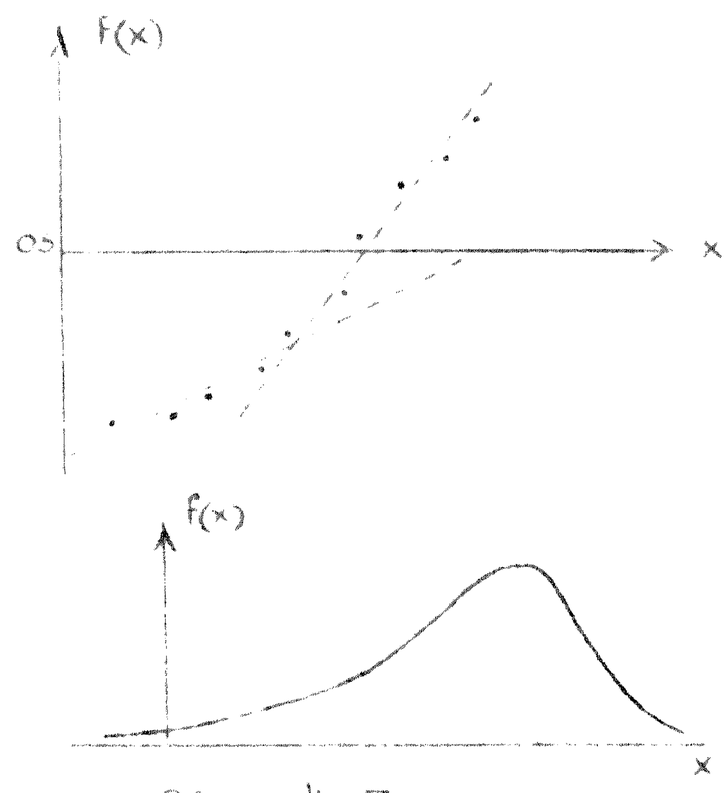


fig. 4.7
Verdeling scheef naar links

Gewoonlijk is de spreiding van de punten om de lijn zo groot, dat alleen sterke afwijkingen van een normale verdeling zichtbaar worden, vooral bij kleine steekproeven. De grafische methode is daarom alleen geschikt om snel een eerste indruk van de verdeling te verkrijgen.

Is de verdeling van $\underline{x}$ scheef, dan kan geprobeerd worden of wellicht log $\underline{x}$ normaal verdeeld is Dit kan alleen wanneer alle x_i positief zijn (te bereiken door bij x_i een constante op te tellen) Met $\underline{x}_i' = \log \underline{x}_i$ kan dan bovenstaande grafische methode uitgevoerd worden. Het is hierbij niet nodig de logarithmes van de x_i werke-

lijk op te zoeken, daar er naast het gewone waarschijnlijkheids-papier met lineaire x-schaal ook logarithmisch waarschijnlijkheids-papier bestaat, waarbij de logarithmische transformatie van de x-schaal reeds is uitgevoerd. Hierop kan men dus weer rechtstreeks de $\underline{x}_i$-waarden en de bijbehorende schattingen $\frac{i-0,3}{n+0,4}$ voor $F(\log x_i)$ aangeven. Voor een meer uitgebreide behandeling van deze methode verwijzen wij naar het boek van AITCHISON and BROWN.[1)]

Tenslotte bekijken wij nog het aanpassen van een exponentiële verdeling voor $\underline{x}$, dus een verdeling waarvoor geldt (vgl. (1.21) van hoofdstuk III):

$$(4.5) \qquad F(x) = \int_0^x \lambda e^{-\lambda v} dv = 1 - e^{-\lambda x}$$

Hierbij kan gewoon logarithmisch papier voor een grafisch onderzoek gebruikt worden, daar

$$(4.6) \qquad z = \log(1-F(x)) = \lambda x$$

is, dus z een lineaire functie van x. Op dit papier kan dan langs de logarithmische schaal een schatting voor $1-F(x)$ dus bijv. $1 - \frac{i-0,3}{n+0,4} = \frac{n-i-0,7}{n+0,4}$ en langs de lineaire schaal x_1 uitgezet worden, waarbij dan de x_i weer naar opklimmende grootte gerangschikt moeten worden. Men kan ook de x_i naar dalende grootte rangschikken en dan $1-F(x)$ schatten door $\frac{i-0,3}{n+0,4}$.

Voorbeeld 4.1 Hoogwaterstanden

In voorbeeld 2.3 van hoofdstuk V zijn wij er van uitgegaan, dat de frequentieverdeling van de hoogwaterstanden voor niet te kleine waarden in goede benadering exponentieel is. Vanaf een bepaalde waarde x_o geldt dan (vgl. (2.17) uit hoofdstuk V)

$$F\left[\underline{x} \geq x \mid \underline{x} \geq x_o\right] = e^{-\lambda(x-x_o)}$$

De juistheid van dit uitgangspunt kan op de volgende wijze onder-

---

1) J. AITCHISON and J.A.C. BROWN, The lognormal distribution, Cambridge University Press (1957)

zocht worden.

Rijkswaterstaat beschikt over de hoogwaterstanden in Hoek van Holland vanaf 1888. Tabel 4 I bevat hiervan een uittreksel, namelijk de standen ≥ 2,42m + N.A.P. in de periode 1888-1956 met de data van optreden. Uit de tabellen van Rijkswaterstaat kunnen allerlei nieuwe tabellen afgeleid worden, bijvoorbeeld tabellen die het aantal overschrijdingen van een bepaald peil x opgeven in de periode 1888-1956. In verband met de onderzoekingen voor de

Tabel 4.I

[illegible]standen te Hoek van Holland

1888-1956

jaar	datum	stand	jaar	datum	stand
1953	1 februari	385	1940	6 december	265
1894	22 december	328	1953	1 februari	265
1916	13 januari	300	1921	6 november	263
1954	23 december	300	1895	23 januari	262
1906	12 maart	297	1912	11 november	262
1904	30 december	296	1946	23 februari	256
1928	26 november	296	1917	2 december	254
1889	9 februari	276	1930	23 november	253
1936	1 december	274	1936	1 december	253
1949	1 maart	270	1897	19 juni	252
1954	24 december	270	1954	22 december	252
1895	7 december	268	1905	7 januari	250
1897	29 november	268	1945	19 januari	246
1943	7 april	268	1917	26 november	244
1944	26 januari	267	1911	30 september	243
1908	23 november	266	1936	18 oktober	242

Delta-commissie interesseerde men zich vooral voor hoge waterstanden in de wintermaanden januari, november en december, die optraden tijdens depressies welke een bepaalde baan volgden. Tabel 4.II bevat in de tweede kolom voor deze hoogwaterstanden de aantallen overschrijdingen in de winters 1888/89 tot en met 1938/39 en 1945/46 tot en met 1956/57 voor de peilen boven 2,42 meter. Daar

gedurende de Tweede Wereldoorlog geen weerkaarten getekend konden worden, was het niet mogelijk de selectie voor de winters van 1939/40 tot en met 1944/45 toe te passen. Het gemiddelde aantal overschrijdingen per jaar van een bepaald peil x verkrijgt men door deze aantallen te delen door het aantal jaren n=63. De overschrijdingskans $1-F(x_i)$ van het peil x_i kan nu voor grote waarden

Tabel 4.II

Hoogste hoogwaterstanden uit naar banen geselecteerde depressies in de winters 1888/89 t/m 1938/39 en 1945/46 t/m 1956/57

H.W.in m	aantal overschrijdingen	$\frac{i-0,3}{63,4}$	H.W.in m	aantal overschrijdingen	$\frac{i-0,3}{63,4}$
3,85	1	0,011	2,54	14	0,216
3,28	2	0,027	2,53	15	0,232
3,00	4	0,043 0,058	2,52	16	0,248
2,96	6	0,074 0,090	2,50	17	0,263
2,74	7	0,106	2,44	18	0,279
2,68	9	0,121 0,137			
2,66	10	0,153			
2,63	11	0,169			
2,62	13	0,184 0,200			

van x_i geschat worden door $\frac{i-0,3}{n+0,4} = \frac{i-0,3}{63+0,4}$, waarin i het nummer aangeeft van de naar afdalende grootte gerangschikte waarnemingen. Voor waarden van x, zoals x=3,00 m ontstaan op deze wijze twee schattingen, aangezien de waterstand van 3,00 m tweemaal is waargenomen en dus in de gerangschikte waarnemingen voorkomt met twee verschillende rangnummers. De schattingen $\frac{i-0,3}{63,4}$ staan in de derde kolom van tabel 4.II.

In figuur 4.8 zijn de punten $(x_i; \frac{i-0,3}{63,4})$ uitgezet, waarbij langs de horizontale as de metrische schaal gebruikt is en langs de verticale de logarithmische. Er zijn meer punten getekend dan in tabel 4.II vermeld werden. Vanaf ongeveer 1,70 m

liggen de punten vrij nauwkeurig op een rechte lijn, zodat de exponentiële verdeling een goede benadering vormt voor de frequentieverdeling van de hoogwaterstanden boven 1,70 m.

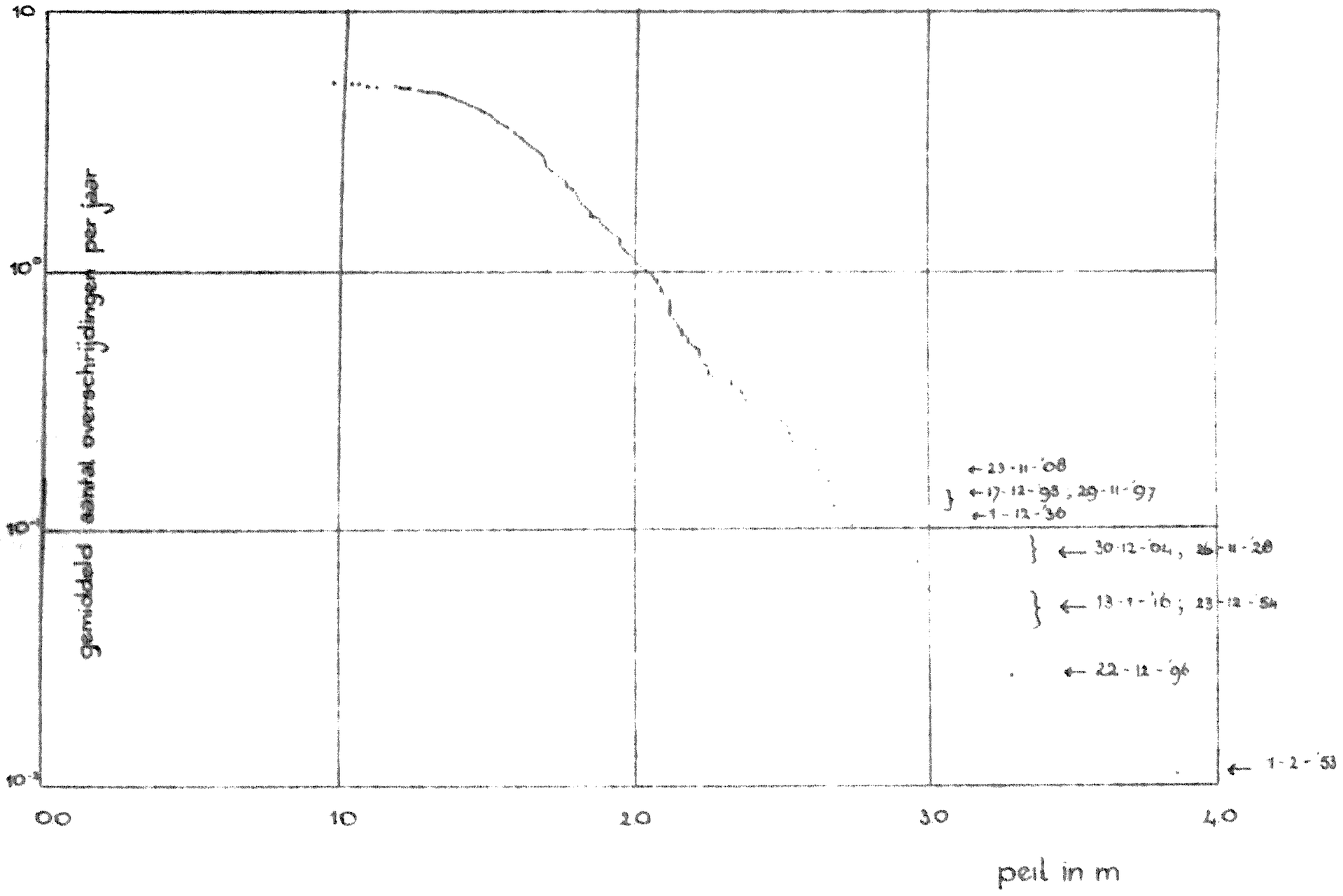


fig. 4.8

Verdeling van hoogwaterstanden te Hoek van Holland

Er bestaat ook dubbel logarithmisch papier (beide schalen logarithmisch), waarmee men dus zou kunnen nagaan of log x exponentieel verdeeld is.

Ook andere transformaties van de x-schaal en ook transformaties van de y-schaal, die op een andere dan de normale of exponentiële verdeling berusten zijn mogelijk om de verdelingsfunctie tot een rechte lijn om te vormen. Hiervoor bestaat echter geen speciaal grafiekenpapier; de transformaties moet men dan dus zelf berekenen, hetgeen de methode zeer bewerkelijk maakt.

Naast grafische methoden bestaan er ook andere methoden om de aanpassing van verdelingen te onderzoeken. Sommige zijn algemeen, dus bij iedere verdeling te gebruiken, andere maken gebruik van bijzondere eigenschappen van de beschouwde verdelingen. Wij behan-

delen alleen een algemene methode nl. de methode van de χ^2-toetsen voor aanpassing.

χ^2-toets voor aanpassing bij een volledig gespecificeerde verdeling

Wil men onderzoeken of een stel waarnemingen, $x_1,\ldots,x_n$, uit een bekende verdeling (bekend wat betreft de vorm èn de waarde van de voorkomende parameters) afkomstig kunnen zijn, dan kan men als volgt te werk gaan.

Verdeel het gebied waarin de steekproefwaarden kunnen liggen in een aantal vakken. Voor ieder vak is dan te berekenen hoe groot de kans is voor een waarneming om in dit vak terecht te komen. Worden r vakken gebruikt, zijn de kansen resp. $p_1, p_2, \ldots, p_r$ (met $\sum_1^r p_i = 1$) en bestaat de steekproef uit n waarnemingen, dan zijn de verwachte aantallen waarnemingen per vak $np_1, np_2, \ldots, np_r$.

In verband met het onderscheidingsvermogen van de toets is het wenselijk de vakken zo te kiezen dat $np_i \approx 12$ bij n=200; $np_i \approx 20$ bij n=400 en $np_i \approx 30$ bij n=1000; dit zijn slechts ruwe richtlijnen. Daar de staarten van de verdeling meestal het belangrijkst zijn, kan men hier kleinere vakken, d.w.z. vakken met kleinere np_i, kiezen, waarbij np_i echter niet kleiner dan 1 mag worden. Soms zijn de waarnemingen alleen als frequenties in intervallen gegeven. In dit geval moet men deze intervallen of combinaties hiervan als vakken nemen.

Noemt men de aantallen waarnemingen uit de steekproef in de r vakken respectievelijk $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_r$, dan geldt:

$$(4.7) \qquad \mathcal{E}\underline{f}_i = np_i \qquad \text{en} \qquad \sum_1^r \underline{f}_i = n.$$

De toetsingsgrootheid is nu

$$(4.8) \qquad \underline{t} = \sum_{i=1}^{r} \frac{(\underline{f}_i - np_i)^r}{np_i} .$$

Onder de nulhypothese dat de kansen voor de vakken werkelijk p_i zijn, heeft $\underline{t}$ bij benadering een χ^2_{r-1}-verdeling (vgl. (1.23) van hoofdstuk III), vandaar de naam χ^2-toets. Geeft de steekproef $x_1,\ldots,x_n$ de aantallen $f_1,\ldots,f_r$, dan kan men met (4.8) de bij deze steekproef behorende waarde t van $\underline{t}$ berekenen. De overschrijdingskans van t, $P[\underline{t} \geq t]$, wordt dan met behulp van de χ^2_{r-1}-verdeling bepaald. Is de nulhypothese niet juist, bezitten de kansen p_i

dus andere waarden dan ondersteld werd, dan zullen de f_i een grotere kans bezitten om sterk van np_i af te wijken dan onder H_o. Dit betekent dan een grotere kans op een grote waarde van t, met dus een kleine rechtséénzijdige overschrijdingskans. Hieruit volgt dat de χ^2-toets steeds rechtséénzijdig gebruikt moet worden.

De benadering van de verdeling van t met een χ^2-verdeling wordt slechter naarmate n en r kleiner zijn, maar is toch altijd nog redelijk goed, zolang $r > 2$ en sommige $np_i > 5$ zijn.

Bij een grote overschrijdingskans zal men dus de hypothese dat $x_1,\ldots,x_n$ uit de getoetste verdeling afkomstig zijn niet verwerpen. Hierbij is nog één opmerking van belang. Met de χ^2-toets wordt nl. in feite alleen getoetst of de vakken de gegeven kansen $p_1,\ldots,p_r$ bezitten. Deze zelfde kansen kunnen echter ook wel van andere verdelingen, dan die getoetst werd, afkomstig zijn. De χ^2-toets maakt geen onderscheid tussen al die verdelingen die de gegeven kansen $p_1,\ldots,p_r$ voor de gekozen vakken opleveren.

In het voorgaande is stilzwijgend ondersteld dat de getoetste verdeling continu was. De χ^2-toets kan echter evengoed toegepast worden bij een discrete verdeling. De discrete waarden die een waarneming voor x nu kan aannemen kunnen hierbij als vakken fungeren. Is n niet groot, dan zullen meerdere naburige waarden in één vak verenigd moeten worden om toch niet te kleine np_i te krijgen vooral aan de staarten.

χ^2-toets voor het aanpassen van een verdeling met onbekende parameters.

Vaker dan de situatie, die wij in het voorgaande beschreven, komt het voor dat men een verdeling wil aanpassen en toetsen, die wel een bekende vorm heeft, maar waarvan de verdelingsfunctie nog van een aantal onbekende parameters afhangt, die ook uit de steekproef geschat moeten worden. De kansen voor de vakken, die wij nu $\pi_1,\ldots,\pi_r$ zullen noemen, hangen dus ook nog van deze onbekende parameters af. De indeling in vakken wordt meestal gekozen met behulp van schattingen voor de onbekende parameters uit de oorspronkelijke waarnemingen $x_1,\ldots,x_n$. Overigens geldt hiervoor hetzelfde als wij in het voorafgaande geval opmerkten.

Om nu de toetsingsgrootheid te berekenen worden voor $\pi_1,\ldots,\pi_r$ schattingen $p_1,\ldots,p_r$ berekend uit schattingen voor de

onbekende parameters. De toetsingsgrootheid is dan weer

$$\underline{t} = \sum \frac{(\underline{f}_i - n\underline{p}_i)^2}{n\underline{p}_i} \quad . \tag{4.9}$$

Voldoen de schattingen voor de onbekende parameters aan bepaalde eisen dan heeft $\underline{t}$ weer bij benadering een χ^2-verdeling, maar nu met r-1-s vrijheidsgraden, als s het aantal onafhankelijke onbekende parameters is. Op de eisen waaraan de schattingen moeten voldoen en de methoden om dergelijke schattingen te vinden kunnen wij hier niet ingaan. Gewoonlijk leiden deze schattingsmethoden tot moeilijk of slechts bij benadering oplosbare vergelijkingen. Men neemt daarom meestal zijn toevlucht tot eenvoudiger schattingen voor de parameters met behulp van de momenten-methode, die in paragraaf 4 van hoofdstuk V besproken werd.

Voorbeeld 4.2 Verkoop van flesjes was

In een warenhuis wil men in verband met het bepalen van een optimale voorraadpolitiek nagaan of de frequentieverdeling van de verkopen per dag van flesjes vloerwas aansluit bij een Poisson-verdeling. Gedurende een reeks weken wordt daartoe iedere dag opgenomen hoeveel flesjes er verkocht worden. Men vindt de volgende resultaten:

Tabel 4.III

Verkopen van flesjes was per dag

aantal flesjes i	frequentie f_i
0	22
1	10
2	7
3	1
4	0
5	0
6	1
alle overige frequenties nul	

De parameter λ van de Poissonverdeling is onbekend en wordt geschat met de momentenmethode. Het gemiddelde in de steekproef bedraagt $\frac{1}{47}(22.0 + 10.1 + 7.2 + 1.3 + 1.6) = 0{,}805$. De verwachting van de Poissonverdeling is $\mathcal{E}\underline{x} = \lambda$ (vgl. tabel 4.I van hoofdstuk III), zodat als schatting $\hat{\lambda}$ van λ gevonden wordt

$$\hat{\lambda} = 0{,}805.$$

Als "vakken" kan men hier kiezen de afzonderlijke waarden 0,1,2 enzovoorts. De tweede kolom van tabel 4.IV geeft de kansen p_i voor $i = 0,1,\ldots,6$.[1] Aangezien de verwachte aantallen voor $i \geq 3$

Tabel 4.IV

Aanpassing van een Poissonverdeling

i	$P[\underline{x}=i \mid \lambda=0{,}805]$	np_i	$\frac{(f_i - np_i)^2}{np_i}$
0	0,4471	18,33	0,735
1	0,3599	14,76	1,535
2	0,1449	5,94	0,189
3	0,0389	1,97	0,005
4	0,0078		
5	0,0013		
6	0,0002		

erg klein zijn, voegen wij de waarden van $i \geq 3$ alle samen tot één vak, zodat wij vier vakken krijgen en $r=4$ is. Kolom 3 van tabel 4.IV bevat de verwachte aantallen np_i voor ieder vak en kolom 4 de waarden $\frac{(f_i - np_i)^2}{np_i}$. De toetsingsgrootheid $\underline{t}$ neemt de waarde

$$\sum_{i=1}^{4} \frac{(f_i - np_i)^2}{np_i} = 2{,}46$$

aan. Deze grootheid bezit bij benadering een χ^2-verdeling met $r-s-1=2$ vrijheidsgraden, daar er één parameter geschat is. De overschrijdingskans bij een χ^2-verdeling met twee vrijheidsgraden

---

1) Deze kansen zijn overgenomen uit: T. KITAGAWA, Tables of the Poisson Distribution, Baifukan (1952), Tokyo.

van de waarde 2,464 bedraagt 0,29 of 29%. Wij zien derhalve dat de aansluiting heel behoorlijk is.

Keuze tussen verschillende verdelingen

De χ^2-toets wordt onderscheidender naar mate men van grotere steekproeven uitgaat. Dit betekent dat ook kleine afwijkingen van H_o kleine overschrijdingskansen geven indien n maar groot genoeg is. Het is daarom moeilijk een grens voor de overschrijdingskans vast te leggen; deze grens, de onbetrouwbaarheidsdrempel α_o, zou afhankelijk van n gekozen moeten worden. Van deze moeilijkheid heeft men minder last indien men een keuze moet doen tussen verschillende mogelijke verdelingen. De χ^2-toets kan dan als kriterium gebruikt worden door de toets op ieder van de verdelingen toe te passen en die verdeling te kiezen, die de grootste overschrijdingskans geeft. Men heeft dan nl. die verdeling uitgezocht waarbij men, althans op grond van de χ^2-toets, het minst geneigd is de hypothese, dat de waarnemingen $x_1,\ldots,x_n$ uit deze verdeling afkomstig zijn, te verwerpen.